Note. "Reaching for the SDGs: the untapped potential of Tanzania's water supply, sanitation and hygiene (WASH) sector." The Water Blog. Word Bank Blogs, 29 Mar. 2018.

AN ANALYSIS OF IMPACT EVALUATIONS OF WATER, SANITATION, AND HYGIENE (WASH) INTERVENTIONS IN RURAL SUB-SAHARAN AFRICA

A REVIEW OF A LITERATURE SAMPLE FROM THE 3IE DEVELOPMENT EVIDENCE PORTAL

R.B. Bader

TU Delft

An Analysis of Impact Evaluations of Water, Sanitation, and Hygiene (WASH) Interventions in Rural Sub-Saharan Africa

A Review of a Literature Sample from the 3ie Development Evidence Portal

By

R. Bader

in partial fulfilment of the requirements for the degree of

Master of Science

in Civil Engineering

at the Delft University of Technology

Project duration: September 08, 2022 – April 17, 2023

Thesis committee: Dr. M. W. Ertsen, TU Delft,

Dr. D. E. Abraham, TU Delft

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

PREFACE

In light of growing recognition of the importance of evidence-based research, implementation, and decision-making in international development, including in the water development sector, this study aims to explore the field of evaluations on foreign aid-funded interventions in the water sector of Sub-Saharan Africa, using the 3ie Development Evidence Portal as a resource to identify relevant studies, analyze the research field and content of those studies, and identify gaps in the literature. The study seeks to provide insights into the characteristics of evaluation studies of water interventions, the configuration of the research community, and gaps in the literature.

This study is relevant to researchers, policymakers, and practitioners seeking to make evidence-based decisions in the sector. It is hoped that the findings of this study will contribute to the design and implementation of effective interventions to improve water security in rural SSA, ultimately enhancing public health, food security, and economic development in the region.

I would like to express my deepest gratitude to my two advisors, Maurits Ertsen and Edo Abraham, for your exceptional guidance and support throughout my thesis work. I truly appreciate your time and effort, and I will always be grateful for the role you played in helping me to achieve this milestone.

I would also like to offer a hue thanks for my mom for your unwavering support throughout my MSc. Your love, encouragement, and guidance have been my rock during the challenging times. I could not have done it without you. You've been my source of strength and inspiration, and I am so fortunate to have you in my life.

Lastly, I would like to thank my truly amazing friends. I couldn't have made it through this crazy journey without you guys. Thanks for being my rock and putting up with my stress-induced meltdowns. You all deserve a trophy for being the most patient and supportive friends a person could ask for!

Delft, April 17, 2023

Rita Bader

Executive Summary

This report presents the findings of a comprehensive analysis of the literature on impact evaluations of foreign aid funded interventions in the water sector of rural Sub-Saharan Africa (SSA). The study was conducted using a dataset extracted from the International Initiative for Impact Evaluation (3ie) Development Evidence Portal (DEP), which is one of the largest repositories in the world for evidence-based research in Low and Middle Income Countries (LMIC). The extracted publications underwent systematic screening based on a set of predetermined criteria. Mainly, studies that pertained to the evaluation of interventions that included infrastructure development and water resource management were solely included, all the while excluding papers that focused on interventions related solely to education, behavioral change, awareness campaigns, or the exclusive provision of soap. After the classification of the initial set of studies, it was found that interventions related to Water, Sanitation and Hygiene (WASH) dominated the dataset, so those related to agricultural productivity, which was the second overarching research sector were excluded. The final dataset amounted to a total of 30 papers. The set of publications was then classified and analyzed using bibliographic and network analyses in order to explore the structure of the research community formed within the dataset. The content of the papers was then analyzed and situated within the external literature, mainly in order to validate and strengthen the conclusions drawn from the analysis, contextualize the findings and recommendations from the study, and identify gaps and areas for further research.

The study reviewed a dataset comprised solely of impact evaluations of randomized controlled trials (RCT's) related to WASH interventions. The trials were spread across nine countries, with the majority being in Kenya, and their duration varied widely, from eight weeks to 29 months. Most of the interventions focused on traditional low-cost household water treatment hardware, with chlorination being the most popular. As for the targeted outcomes, the majority of the trials investigated health-related outcomes for children under five years, with diarrhea prevalence being the most recurrent outcome studied. The results further showed significant heterogeneity among the trial findings, especially regarding diarrhea prevalence, suggesting that intervention effectiveness may depend on various conditions that research has not fully explained yet.

The research community within the dataset showed disparities in the distribution of trials and studies across SSA, with the majority of studies conducted by organizations and institutions from the Global North and published in journals mainly oriented towards the Global North audience. Furthermore, the studies employed similar approaches to investigate highly recurrent outcomes, further suggesting a potential lack of diversity in the knowledge pool. The dataset focused on low-cost interventions that could be implemented on a wide scale and showed a preference for RCT's reported to be chosen due to their ability to establish causal relationships between interventions and outcomes, minimize bias and produce generalizable findings. However, the RCT's in the studies included recurrent design limitations that compromised their promise of minimizing bias, including their inability of blinding and their reliance on subjective outcome indicators. Furthermore, the RCT's were found to insufficiently report on data that would aid in the generalizability of their findings, including the challenges experienced in the field and the logistical information of the trials. The study therefore highlighted the need for greater diversity in approaches and

knowledge, local representation, and greater innovation and experimentation in research on WASH interventions in SSA.

To address these areas for improvement, the study makes several recommendations. First, this study emphasizes the need for greater representation of organizations and research output from the Global South, and the increase in their collaborations with the Global North, as these ties proved to generate high-impact research. Second, the development of standardized reporting guidelines and data sharing protocols to improve the transparency and reproducibility of impact evaluations is recommended. Third, a greater diversity in the targeted outcomes and implemented approaches and the upgrading of the quality of the implemented research was also commended.

All in all, this study provides important insights into the state of impact evaluations in LMICs, highlighting both the progress that has been made and the challenges that remain. By providing recommendations for improving the quality and impact of impact evaluations in LMICs, this study has the potential to inform the work of researchers, policymakers, and practitioners working to improve development outcomes in these countries.

Glossary

LIST OF ACRONYMS

3ie	International Initiative for Impact Evaluation
ARI	Acute Respiratory Infection
BAC	Before and After Concurrent Control
CFU	Colony-Forming Unit
CLTS	Community-led total sanitation
CNCI	Category Normalized Citation Impact
DEP	Development Evidence Portal
DRC	Democratic Republic of Congo
E. Coli	Escherichia Coli
EED	Enteric Dysfunction
IYCF	Infant and Young Child Feeding
JCI	Journal Citation Indicator
JIF	Journal Impact Factor
LMIC	Low- and Middle-Income Countries
MDG	Millennium Development Goal
NP	Number of Publications
NTU	Nephelometric Turbidity Units
OECD	Organization for Economic Co-operation and Development
PLA	Participatory Learning and Action (PLA)

PET	Polyethylene Terephthalate
PHAST	Participatory Hygiene and Sanitation Transformation (PHAST)
POU	Point of Use
POS	Point of Source
RCT	Randomized Controlled Trial
SDG	Sustainable Development Goal
SHINE	Sanitation Hygiene Infant Nutrition Efficacy
SODIS	Solar disinfection
SSA	Sub-Saharan Africa
UK	United Kingdom
UN	United Nations
UNDP	United Nations Development Program
UNICEF	United Nations Children's Fund
USA	United States of America
VEA	Villages et Ecoles Assainis
WASH	Water, Sanitation and Hygiene
WBG	World Bank Group
WHO	World Health Organization
WoS	Web of Science

1
Introduction

The lack of water security in rural Sub-Saharan Africa (SSA) has far-reaching implications for the well-being and development of individuals and communities in the region. According to the United Nations (UN), more than 40% of the population in Sub-Saharan Africa lacks access to safe drinking water, and approximately 70% of the region's population relies on unimproved sanitation facilities (UN, 2019). As a result, waterborne diseases such as diarrhea, cholera, and typhoid fever are prevalent, and children under five years of age are particularly vulnerable (Adeyemo, 2020). Furthermore, and according to the World Health Organization (WHO), around 115 people in rural SSA die every hour from diseases linked to poor sanitation, poor hygiene practices, and unsafe drinking (WHO, 2019).

In addition to health risks, water insecurity also affects agriculture, which is the primary livelihood for many rural households in SSA. Lack of access to water and irrigation facilities leads to crop failures and low agricultural productivity, which, in turn, impacts food security and ultimately contributes to poverty in rural communities (Adeyemo, 2020). Concurrently, water insecurity hinders economic growth and development in rural areas, as industries and businesses cannot operate without adequate water supply. A report by the World Bank Group (WBG) estimates that water scarcity costs the whole of SSA 1% to 2% of its GDP every year, with the agriculture sector being the hardest hit due to the lack of irrigation systems and water storage facilities (WBG, 2020).

Water insecurity in rural SSA has a disproportionate impact on vulnerable communities, such as older people, women, and people with disabilities. Older people may experience greater difficulty in accessing water sources due to physical limitations and limited mobility, making it challenging for them to collect water from distant sources. People with disabilities, including physical disabilities and visual impairments, may also face significant barriers in accessing water sources, exacerbating their vulnerability to water insecurity. Women in particular bear the brunt of water insecurity as they spend significant time and energy collecting water, and may face additional challenges due to a lack of safe and secure water sources, leading to increased risks of violence and sexual harassment. According to UNICEF, women and girls in SSA spend 40 billion hours every year collecting water, which is equivalent to a year's worth of labor by the entire workforce of France. This limits their opportunities for education, or economic and community participation, and considerably affects their health (UNICEF, 2016).

Furthermore, water insecurity has intergenerational effects that create a vicious self-actualizing cycle of poverty, ill health, and limited opportunities. Children born into households without access to safe water and sanitation are more likely to suffer from waterborne diseases, malnutrition, and stunting, which in turn can lead to impaired cognitive development and poor academic performance (UNICEF, 2019). The lack of access to safe water and sanitation facilities also limits the time that women and children can spend on other productive activities (UNDP, 2020). Accordingly, addressing water insecurity in rural SSA may be crucial to break this cycle through sustainable multifaceted solutions that provide access to safe water and sanitation facilities.

Concurrently, in recent years, there has been a growing recognition of the importance of evidence-based research, implementation, and decision-making, including in the water development sector. This shift reflects a broader trend towards evidence-based approaches in international development, which emphasizes the need for rigorous evaluation of interventions and the use of data and evidence to guide policy and practice (Bovens et al., 2016). The shift is largely driven by the recognition that past approaches have not always been entirely effective in addressing the complex challenges of ensuring water security for communities. Interventions in the water sector have been reported to have failed in achieving their desired outcomes, and in some cases, have even led to unintended consequences such as social and environmental harms (Baker et al., 2019; Foster et al., 2018; Cissé et al., 2016).

In addition, the increasing focus on accountability and transparency in the aid sector has led to a demand for rigorous evaluation of interventions and evidence-based decision-making (Kenny et al., 2019). Funders of water development projects, such as governments and multilateral organizations, have also recognized the importance of evidence-based approaches in achieving sustainable and long-lasting impact (OECD, 2019). Likewise, the adoption of the UN's Sustainable Development Goal (SDG), particularly Goal 6 which aims to ensure availability and sustainable management of water and sanitation for all, has reinforced the need for evidence-based approaches to achieving water security (United Nations, 2015).

One key organization driving evidence-based research is the International Initiative for Impact Evaluation (3ie), which operates a database of impact evaluations and systematic reviews in the development research field, including the water supply and sanitation sector (3ie, n.d.). This database has the potential to provide a valuable resource for researchers, policymakers, and practitioners seeking to make evidence-based decisions in the sector.

Accordingly, this study aims to explore the field of evaluations on foreign aid funded interventions in the water sector of SSA. The 3ie Development Evidence Portal (DEP) serves as a resource to identify relevant studies, analyze the research field and content of those studies, and identify gaps in the literature.

In particular, the study aims to answer the following questions:

1. What are the different characteristics of the evaluation studies of the water interventions in the region?
2. How is the research community of the dataset evaluations configured?
3. What gaps in the literature can be identified?
4. Does the dataset, its different characteristics and research community mirror those of the general literature?

To answer these questions, the study starts with identifying relevant publications in the 3ie DEP pertaining to the water sector in rural SSA based on systematic inclusion criteria. The study goes on to extract the characteristics data of the amassed literature, and perform bibliometric and network analyses for an insight into the research community of the dataset. The study continues by analyzing the results from the characterization, the bibliometric and network analyses and the content of the publications, all the while mirroring those findings with the general literature to identify gaps within the field and finally present some recommendations for future research.

At large, the findings of this study contributes to the body of knowledge related to the water foreign aid sector and to the optimization of its effectiveness and advancement. By using the 3ie DEP, the study utilizes and

analyzes the products of a widely used and reputable source in the international development field. This approach aspires to draw findings that are relevant to policymakers, funders, practitioners, and researchers working in the field. Furthermore, by analyzing the research community of the field, the study identifies areas where the community has made significant contributions and areas where more research, collaboration, or innovative and diverse approaches and practices are needed. Additionally, by comparing the findings from the 3ie evidence portal to the general literature, the study assesses whether the selected publications mirror the broader research landscape and provides an insight into the current state of research on the water sector in rural SSA and potential literature gaps in it. It also assesses the degree of comparability, and by extension, the generalizability of its findings. The study ultimately informs policy and practice by providing evidence-based recommendations for the design and implementation of water interventions along with their evaluations in rural SSA. These gaps could inform future research, aid interventions and help ensure that limited resources are allocated effectively.

The remaining of this report is organized as follows. Chapter 2 begins by providing a brief overview onto the 3ie Development Evidence Portal (DEP), and proceeds to illustrate the methodology adopted to identify and classify the publications from the portal, conduct the bibliometric and network analyses and analyze the finding. Chapter 3 presents the results of the publications mining and classification processes, the bibliometric and network analyses, as well as the findings of exploring the content of the dataset. Chapter 4 was specifically dedicated for the dominant evaluation process design in the dataset, the randomized controlled trial design. Chapter 5 groups all of the findings together and analyzes their implications, followed by the Chapter 6, where the general conclusions of the study are made and the drawn recommendations are presented.

2
Methodology

The following chapter details the steps adopted to find the initial body of literature and the subsequent selection of the appropriate studies to be eventually classified and analyzed. The first part of this chapter starts off with a brief overview of the 3ie Development Portal, the database from which the publications considered in this study were extracted. Next up, the inclusion and exclusion criteria employed to refine the set of data obtained in the first step are relayed. The second part of the chapter details the steps performed in the bibliometric and network analyses done on the finalized dataset in order to gain insight on the research community structure of the dataset. The chapter ends with a detailing of the methodology adopted in the analysis of the content of the dataset and its representativeness with respect to the general literature.

2.1 PUBLICATIONS SEARCH METHODOLOGY

2.1.1 OVERVIEW OF THE 3IE DEVELOPMENT EVIDENCE PORTAL

The 3ie Development Evidence Portal (DEP) is an online repository that provides access to evaluations of international development interventions, with a focus on low- and middle-income countries. The portal is offered by the International Initiative for Impact Evaluation (3ie), which is an independent organization established in 2008, that generates and promotes the use of rigorous evidence in development policy and practice (3ie, n.d.).

3ie is funded by various development agencies and foundations, and has a mandate to support the generation and use of evidence to inform development policy and practice worldwide. The 3ie portal contains a wide range of evaluations, including impact evaluations, systematic reviews, and meta-analyses. In addition to evaluations commissioned by 3ie, the portal also includes evaluations from other organizations, attempting to warrant that the portal provides a comprehensive overview of the available evidence on development interventions. (3ie DEP, n.d.). The evaluations cover a wide range of development sectors, including health, education, agriculture, and governance. The portal also provides tools for searching and browsing evaluations along with an indication of the quality of the content, making it a valuable resource for development practitioners, policymakers, and researchers.

By browsing through the database, three overarching sectors were indicated by the portal to be closely relevant to international development in the water sector: water and sanitation, agriculture and health. In the water and agricultural sectors, some of the most important organizations and researchers contributing to the 3ie Development Evidence Portal include:

- The WBG

- The International Water Management Institute
- The International Centre for Integrated Mountain Development
- The International Development Research Centre
- The Stockholm International Water Institute
- The University of California, Berkeley
- The Institute of Development Studies

In the health sector, some of the most important organizations and researchers contributing to the 3ie Development Evidence Portal include:

- The World Health Organization (WHO)
- The WBG
- The Bill and Melinda Gates Foundation
- The University of California, Berkeley
- The London School of Hygiene and Tropical Medicine
- The International Food Policy Research Institute
- The World Vision International

It is worth noting that this list is not exhaustive, as 3ie Development Evidence Portal is open for contributions from many other organizations and researchers (3ie DEP, 2023).

2.1.2 KEYWORDS AND SELECTION CRITERIA

The finalized dataset that this study revolved around was obtained by applying thorough data mining from the 3ie portal, followed by a screening and the refinement of the resultant set of publications. These steps are elaborated upon in the following:

INITIAL PUBLICATIONS MINING

The keywords employed in this process were determined in an iterative way until an adequate set of studies was attained.

1) The search started with a generic search using the keywords “water” and “rural”, which produced over 747 results, of which 234 belonged to the “Health” sector, 210 to the “Water, sanitation and waste management”, 144 to the “Agriculture, fishing and forestry”, 72 to “Social protection”.
    a. The bulk of the studies was aggregated between the Health, Water and Ecosystems sectors, with the first encompassing the largest amount. After browsing through the content of the results and

their database sector-classification, the latter was found to be inconsistent. The inconsistency was due to the fact that the studies that overlapped between two or more sectors did not have clear sector-classification indicators and were hence categorized in only one of their identified sectors.

b. This initial search however offered a broad scope on the content of the database and permitted to identify a number of key and highly relevant studies. By checking if those studies appear in the upcoming more narrowed searches, the accuracy of the results would be tested.

2) The next step was to restrict the search by confining the results to solely include those that had the terms "water" and "rural" in the keywords of the study, but this generated a minuscule number of results (25 hits) and eliminated important articles, so the search was broadened to include "water" and "rural" in the title or abstract.
    a. This strategy would ensure a higher level of accuracy as the focus areas of the results are bound to be related to the Water field if the term was featured in the title, and to a lesser extent, in the abstract.
    b. The search offered 53 results. However, browsing through the results revealed that a considerable amount of the representative articles identified in the step 1.b) were missing. This observation indicated that the search strategy is non-comprehensive.

3) The third step was to widen the search while prompting more accuracy.
    a. Therefore, the synonyms of water and rural were first obtained from Thesaurus. These amounted to the following: "precipitation", "rainwater", "rainfall" and "rain" for "water", along with "pastoral", "farm*", "agrarian", "remote area*", "village" and "countryside" for "rural".
    b. Furthermore, key targeted terms were identified by browsing the themes related to the water sector in the databases of the World Bank Group (WBG Water, n.d.), the Organization for Economic Co-operation and Development (OECD Water, n.d.) and United Nations Water (UN Water, n.d.), as these entities are the world's leading developmental aid institutions.
    c. All three included the terms sanitation, hygiene, WASH (Water, sanitation and hygiene), irrigation, transboundary and wastewater as sub-sectors or indicators, along with composite terms such as water management, water quality, water stress.
    d. Therefore, and since the composite terms already include "water", the new search entry was indicated to identify in the title or abstract the following keywords: water, sanitation, hygiene, WASH, irrigation, transboundary or wastewater. Furthermore, the terms "drought", "flood*", "dam*" and "hydropower" were noted to be recurrent in the preliminary overview of the topic performed at the very beginning and so were also included as targeted terms to be identified in the title or abstract. If the indicative studies identified in the first step were next looked for in the results and were found to be included, then the search results were deemed to be adequate.

SCREENING

The obtained publications were then screened by title and abstract and duplicates were removed. Three exclusion criteria were adopted:

1) studies that did not evaluate the impact of foreign aid as some were found to denote field observations that were made for the purpose of advising future interventions;
2) studies that assessed an intervention related solely to education, behavioral change, awareness campaigns, or the exclusive provision of soap were eliminated. The latter is because the review focuses on a narrowed aspect of water sector interventions, mainly infrastructure development and water resource management;
3) duplicated publications

Since one of the objectives of this study is to analyze the relevant literature-set that could be offered by the 3ie Development Portal and to track the temporal evolution of the themes and methods used in the assessments of water interventions performance, no confining date range was set for the obtained results.

Once the set of data was obtained, the initial step was to differentiate the general sectors of the publications, in terms of them being in the WASH, agricultural or socio-economic sectors. Choosing these sector categories was due to the fact that while browsing the themes related to the water sector in the databases of the WBG, OECD and UN Water, these three sectors were found as overarching sector groups for international development in the water field.

2.2 DATA COLLECTION METHODOLOGY

The next step was to extract information pertaining to the characteristics of the evaluations, which were tabulated into a standardized Microsoft Excel data abstraction form, included in Appendix B.

In general, WASH interventions can be categorized into two main types: supply-side and demand-side interventions. Supply-side interventions are those that aim to improve the availability and quality of WASH facilities and services, such as the construction of water supply systems or the installation of sanitation facilities. Demand-side interventions, on the other hand, focus on changing the behaviors and attitudes of individuals and communities towards WASH practices, such as promoting handwashing or providing education on safe water storage (UNICEF, 2016).

The interventions were hence classified as being demand or supply-side focused and the following information was also extracted and grouped according to the following categories: country; trial period; type of intervention; targeted outcomes; as well as technical, health and behavioral indicators.

2.3 BIBLIOMETRIC AND NETWORK ANALYSES

Insightful bibliometric and network analyses are able to provide data-driven insights into the structure and impact of the publications, as well as their connections to each other. For context, a bibliometric analysis involves the quantitative analysis of publication and citation data to measure the impact and relative standing of individual publications, authors, or research fields (Glanzel & Schoepflin, 1999). A network analysis, on the other hand, involves the visualization and analysis of the connections and relationships between publications, authors, or research fields. Together, these tools provide insights into the structure and development of research fields no matter their scale, help to identify gaps in knowledge, and support evidence-based (or data-based) decision making (Newman, 2003).

2.3.1 BIBLIOMETRIC ANALYSIS

The bibliometric analysis was based on the bibliographic data of the studies in the dataset. The data was extracted and stratified between three levels: dataset, publishing journals and contributing authors. Accordingly, for each of these three constituents, the temporal productivity and performance were obtained. However, and since none of these indicators can be viewed as being accurately representative on their own, these parameters are meant to be complementary and together provide insight onto the constituents.

DEFINITIONS

The definitions of the bibliometric indicators and their utilized indicators are detailed in what follows:

I. PRODUCTIVITY OF THE CONSTITUENTS

Productivity can be thought of as being represented by publications. Constituent productivity refers to the number of publications produced over a certain period of time. It is often used as a measure of a constituent's research output and can be a factor in determining their academic or professional standing. In this study, constituent productivity is measured by amounting the number of publications produced yearly over the period of the dataset (Lathabai et al., 2017).

II. PERFORMANCE OF THE CONSTITUENTS

As for the performance of the three constituents, their research output was appraised using academic performance indicators.

A. Performance of the Dataset: For the dataset as a whole, the number of citations it received and references it made were utilized. Indeed, by comparing the citations and references of a dataset of publications, some insight into its quality as a whole can be provided (Papaioannou, 2018).

   Citations may be indicative because they can denote that other researchers have found the work to be valuable and have used it as a basis for their own research. High citation counts can suggest that the publication has made a significant contribution to the field and has had a lasting impact (Ioannidis, 2005).

   References may also be important because they provide context for the work and demonstrate the author's familiarity with the relevant literature (Harzing, 2017). A publication with a comprehensive and well-researched list of references can suggest that the author has a strong understanding of the field and has conducted a thorough review of the existing literature (Molina-Montes et al., 2018).

B. Performance of the Journals: For journals, their general performance can be understood through the citation analysis of the journals. Citation analysis is a method of evaluating the influence and impact of published research by analyzing the number and frequency of citations that a given journal receives from other scholarly works.

   Citation analysis, using common indicators, like the Journal Impact Factor (JIF), must control for variables such as subject field, document type, and year of publication in order to be valid in evaluating the quality status of published research. This requirement for journal evaluation is met by the new Journal Citation

Indicator (JCI) developed by Clarivate of the Web of Science (WoS) platform, which provides one number that incorporates the unique attributes of different fields and their publications. The calculations behind JCI are reported to be intricate, necessitating significant computing power, while the end result is straightforward: a single value that is simple to interpret and compare, complementing current journal metrics and supporting responsible use (Shultz & Vaughn, 2018).

The JCI uses another Clarivate measure in its calculation for a given journal: Category Normalized Citation Impact (CNCI). The JCI value is the average CNCI for all articles and reviews published in a journal in the previous three years. It also mitigates for the discrepancy that may result from publications from different fields, document types (articles, reviews, etc.), and year of publication. The resulting number represents a paper's relative citation impact as a ratio of citations compared to a global baseline. A value of 1.0 represents the global average, with values greater than 1.0 indicating greater-than-average citation impact (2.0 being twice the average) and values less than 1.0 indicating less-than-average citation impact (Bornmann, 2020). All JCI values featured in this study are the values for the year 2022.

C. Performance of the Authors: For the authors, the 20 most productive researchers in the dataset were closely investigated. To evaluate their academic impact, their H-index was considered. The H-index is a widely used bibliometric index that attempts to measure both the productivity and impact of an individual researcher's work. Since scholars do not seem to have officially delimited H-index values, and after browsing through multiple articles discussing the matter (Ain et al., 2019; Pluskiewicz et al., 2019; Gagolewski & Mesiar, 2012), the recurrent consensus is that an H-index of 20 is good, 40 is great and 60 is excellent, and therefore those values were used as delineators of different performance levels.

While the H-index may be a practical tool for comparing researchers within a particular field, it ought to be handled with great apprehension as it has also been subject to several criticisms. Some of the critiques of the H-index include:

1. **Self-citation bias**: it can be influenced by the extent to which a researcher cites their own work. Self-citations can inflate a researcher's H-index, and this can create a distorted view of their research impact.

2. **Field-specificity**: it can be influenced by the norms and publication practices of different academic fields. This means that the H-index may not be comparable across different fields of research (Hirsch, 2005).

3. **Time-dependent**: it is a measure that is highly sensitive to the passage of time. Newer researchers may have a lower H-index than more established researchers, even if their work is highly impactful.

4. **Co-authorship bias**: it does not differentiate between a single-author publication and a multi-author publication. As a result, researchers who collaborate frequently may have a higher H-index than those who work independently.

5. **Citation behavior**: it is based on the number of citations a researcher's work receives, but citation behavior can be influenced by a variety of factors, including the popularity of the topic, the quality of the research, and the prestige of the journal in which it is published.

6. **Limited scope**: it only considers publications that have been indexed in certain citation databases, in this case WoS, which may not include all relevant publications in a given field (Costas & Bordons, 2007).

7. **Disserving authors from developing countries**: authors from LMIC's may have fewer opportunities to publish their work in high-impact journals and conferences, which can limit their H-index. In addition, many local publication outlets in developing countries may not be indexed in the databases that are used to calculate the H-index, so the impact of this work may not be fully captured in an author's H-index. They also may have limited access to research resources such as funding, equipment, and supportive research environments, which can make it more difficult for them to produce high-impact research. Lastly, some authors from developing countries may face language barriers that make it difficult for them to publish their work in English-language journals, which are typically given more weight in H-index calculations (Glänzel, 2006).

While there are several other metrics available that are able to infer the academic standing of an author, and despite the criticisms revolving around it, the H-index was used for several reasons:

1. **Focus on impact**: it takes into account the impact of an author's work and measures both the number of publications an author has produced and the number of citations those publications have received, indicating the influence of the author's work on the field.

2. **Works well within same or intertwined fields**: The fields of water supply and sanitation interventions and their impact on health improvement are generally considered to be closely related research fields. There is a growing body of evidence linking access to clean water to improved health outcomes, particularly in low- and middle-income countries (LMICs) where these resources are often lacking (Prüss-Ustün et al., 2019).

   Given the close relationship between WASH and health research, it is likely that the H-index would be a useful metric for assessing the impact of researchers in these fields. While the H-index is generally considered to work best within a single research field, it can also be warranted for evaluating researchers across related fields that share common themes or research questions (Liu et al., 2017). In the case of WASH and health research, both fields are focused on understanding the impact of environmental factors on human health, and researchers in both fields are likely to be interested in similar questions related to the effectiveness of WASH interventions and their impact on health outcomes.

3. **Standardization**: The indicator provides a standardized measure that allows for easy comparison of authors working in the same field or discipline. This can be helpful for identifying experts in a particular area or for assessing the relative impact of different authors.

4. **Transparency**: The H-index is a transparent and objective metric that is widely used in academic circles. It is calculated based on data that is publicly available and can be easily verified.

5. **Widely recognized**: The H-index is a widely recognized metric, making it a useful tool for evaluating authors across different institutions (Costas & Bordons, 2007). So far, there has been no other alternative metric that the general literature has widely agreed upon so the H-index will be used with careful consideration in this study.

METHEDOLOGY

To extract all of the aforementioned information, the Web of Science (WoS) was used. WoS is a citation database and research tool that provides access to a vast collection of scholarly literature across multiple disciplines. It is owned and operated by Clarivate Analytics, and contains bibliographic information, abstracts, and citation data for millions of articles, conference proceedings, and other scholarly materials from over 250 disciplines (Bornmann and Leydesdorff, 2013). One of the key benefits of using WoS is its ability to help researchers track the impact and influence of researchers. Indeed, according to a study by Shultz and Vaughn (2018), WoS is particularly useful for tracking citations because it offers a comprehensive view of the research landscape across multiple disciplines, and allows users to search for and analyze citation data in a variety of ways. Therefore, benefitting from the database's features, the publications in the dataset were identified on it and compiled into a list.

Generating a list in WoS of the publications featured in the finalized dataset extracted from 3ie permitted the easy extraction of all of two main bibliometric indicators of the the constituents of the publications, these constituents consisting of the dataset as a whole entity, the authors and the journals.

2.3.2 NETWORK ANALYSIS

Obtaining a comprehensive representation of a research field can prove to be a difficult endeavor, especially since domineering subjectivity is too easily attainable. To this end, a network analysis was performed to connect the literature across authors, topics, and fields and dissect the structure of the research community of the dataset.

Once again, network analysis is a technique used to study the relationships and patterns of collaboration and communication between researchers, institutions, and scientific disciplines within a field of study. It involves visualizing and analyzing the connections between the different constituents to understand how scientific knowledge is produced, disseminated, and cited. It can provide valuable insights into the structure of scientific collaboration and communication, the interdisciplinary nature of research, and the emergence of new subjects or interests within fields of study. It is commonly used alongside bibliometric studies to map the distribution of research activities and evaluate the impact of scientific knowledge production (Borgatti et al., 2009).

In the context of this study, the analysis was used to examine the academic connections formed within the small knowledge community developed by the data collected from the DEP pertaining to the water and health fields in the rural Sub-Saharan context. Specifically, the network was useful for mapping the scope of the disciplines and their structure, detecting the most influential papers in the field, and identifying key research clusters.

From here onwards, for consistency, the following terminology was employed from here onwards: the terms "local" or "internal" referred to publications within the dataset that is being analyzed in this study. On the other hand, the term "global" or “external” were used to indicate the entire network of publications and citations that exist beyond this subset, including all of the publications and citations that are not represented in the "local" or "internal" dataset.

SOFTWARE TOOL USED

The relationship prevailing amongst the publications of the dataset was analyzed using a software called VOSViewer. The software is a bibliometric tool designed to visualize and analyze scientific networks and research landscapes. It is widely used and has become a popular choice among researchers for several reasons:

1. Firstly, VOSViewer is highly customizable, allowing users to tailor the appearance and functionality of their network visualizations to fit their specific needs;
2. Secondly, the software is user-friendly, making it accessible to a wide range of users with varying levels of expertise;
3. Additionally, VOSViewer provides a number of advanced analysis options, including centrality measures and cluster detection, which enables users to uncover deeper insights into the structure and dynamics of their research networks.

These features, combined with its affordability and ease of use, make VOSViewer a popular choice for network analysis in compared to other software options, which is why it was used in this study (Van Eck, 2010).

To uploaded the data onto VOSViewer, the publications along with their cited references were first exported from the WoS platform in the form of plain text and then upload onto VOSViewer. The software was utilized to conduct two forms of analyses: the citation network and keyword network analyses. The procedure executed, including the VOSViewer features that permitted these analyses, are elaborated upon in what follows.

A. CITATION NETWORK ANALYSIS

The software was used to analyze the citation network within the dataset. By visualizing the connections between publications based on their citation relationships, it was possible to isolate influential publications that are highly cited, as well as clusters of related research.

In particular, analyzing a dataset where the papers cited by each other, those that were not cited by anyone, and the citation network between the papers provided valuable information about the research landscape of a particular field. Studying the papers cited by each other can reveal clusters of closely related research, highlighting the most influential and impactful papers.

Conversely, analyzing the papers that were not cited by anyone can provide insights into neglected areas of research or emerging topics that have not yet gained widespread attention. Finally, examining the citation network between papers can reveal the most important and influential papers in a field, as well as identify potential collaborations and research trends (Elsevier, n.d.).

VOSViewer VISUALIZATION FEATURES

If two papers are close together in a citation map in VOSViewer, it typically means that there is a strong citation relationship between them. This means that the two papers are often cited together, either by the same authors or by different authors. The closer the two papers are to each other in the citation map, the stronger the citation relationship between them is likely to be.

As for the nodes themselves, they correspond to the citation impact of a paper. Papers that are cited more frequently and have thus may have higher influence within their field will typically be represented by larger nodes in the citation map. This was useful for identifying key papers and authors within the network.

The colors of the clusters in a citation map represent different citation groups or clusters of papers that have similar citation patterns. Papers that belong to the same cluster will have the same color in the citation map, making it easy to visualize the different citation groups within the network. The number of clusters and their size are determined by the resolution value. A higher resolution value will result in a larger number of smaller clusters, while a lower resolution value will result in fewer but larger clusters.

By default, VOSViewer uses a resolution value of 1.0, which tends to result in a moderate number of clusters of varying sizes. To obtain an optimized number of clusters that would be insightful enough to allow the identification of small research communities within the dataset and their interactions, the resolution value was iteratively set for 4 or 5 times. At each run, the value was increased and the resulting network was checked for meaningful clustering.

B. KEYWORD NETWORK ANALYSIS

Thematic analysis is a method used in content analysis to identify and analyze recurring themes or patterns within a given dataset. To perform the analysis, a keyword co-occurrence breakdown is performed via VOSViewer to identify trends and relationships between keywords. A visual representation of the interconnections between keywords is therefore visualized by a map, where each node represents a keyword, and the edges between nodes represent the co-occurrence of keywords in the dataset. Furthermore, if two keywords are close to each other, it means that they are frequently mentioned together in the analyzed set of documents. As for the size of the nodes, they are used to represent the frequency of the keywords across the dataset.

Furthermore, VOSViewer is able to generate a parameter termed the total link strength of a keyword, which is a measure of how closely a keyword is related to all other keywords in the map, based on the frequency and strength of their co-occurrence relationships. Therefore, keywords with higher total link strength values are likely to be more important or central to the analyzed topic than keywords with lower values.

For generating an insightful thematic map, the 50 most recurrent keywords were included in a keyword co-occurrence map. From the map, thematic clusters were generated that would represent certain dominant topics or themes within the dataset. To do so, the resolution value based on which the number of clusters is determined was incrementally increased from its default value of 1 until the generated clusters stopped being insightful. Finally, the total link strength values of the top 10 most recurrent keywords were reported.

2.3 REPRESENTATIVENESS AND CONTEXTUALIZATION OF THE DATASET

In what follows, the motives behind conducting a dataset representativeness analysis are detailed, followed by the two main approaches adopted to perform the analysis: the bibliometric representativeness analysis and the data comparability analysis.

A. MOTIVES FOR THE REPRESENTATIVENESS AND CONTEXTUALIZATION ANALYSES

The dataset was evaluated to check if it is representative of the general literature for several reasons. Evaluating whether the dataset is comparable to the general literature can help to assess the generalizability of its findings to other contexts and identify and literature gaps present. This is especially important since the papers considered in this study are all extracted from a single database, the 3ie DEP. Working with a dataset that includes studies extracted solely from the 3ie DEP can limit the scope of the study in several ways. These motives are:

1. **Representativeness of the outcomes of the sample**: The content of the portal may not be a representative sample of all development interventions globally. This means that the studies in the dataset may not reflect the full range of interventions or the full range of outcomes that have been evaluated in the development field.
   This is because the DEP relies on the availability of impact evaluations and other rigorous evaluations, which are often conducted by well-funded organizations or institutions. This means that the DEP may not capture evaluations that were conducted by smaller organizations, or evaluations that did not meet the rigorous standards required for inclusion in the DEP. This may ultimately be indicative of the formation of a closely interconnected research community which limits the scope of the study.
2. **Publication bias**: The studies in the DEP may suffer from publication bias, as negative results are less likely to be published than positive results. This means that the studies in the dataset may over-represent the effectiveness of interventions and underestimate potential negative effects (Petticrew & Roberts, 2006).

To do all of the aforementioned, the degree of representativeness of the dataset was first assessed based on its bibliometric configuration in order to situate the miniature research community formed in the study with respect to the global literature community. This was complemented by the analysis of the different elements of the dataset content in order to contextualize the dataset within the general literature. This was done by either comparing the findings of the papers to those featured in external literature, or by identifying the input that external literature can contribute in light of the dataset content and findings.

B. METHODOLOGY OF THE ANALYSES

Bibliometric Representativeness of the Dataset

In order to put the dataset of this study into perspective, its bibliometric representativeness was first assessed. To do so, its different characteristics were compared with the global literature related to the topic. The WoS platform was used as an insight into the global literature.

Accordingly, the same keywords used to assemble the dataset in the 3ie portal are used. Furthermore, since the 3ie DEP was by nature a repository of development evaluation research, new terms pertaining to academic assessments literature were added to the search in WoS. These terms were (efficiency OR effectiveness OR consequence OR

outcome OR impact OR evaluation OR evaluating OR assessment OR assessing OR effect*OR lesson OR success OR efficacy) were additionally added in the search to be featured either in the title, abstract or article keywords. As for confining the results to English publications of the SSA region, the country and language filters in WoS were used.

The bibliographic data of the DEP dataset and those from the assembled WoS dataset were then compared. In particular, the data consisted of the top 10 most recurrent constituents of the following characteristics: sector categories of the topics, authors, author affiliations, journal name and publishing country. Likewise, the top 10 most internally cited articles within the dataset and the top 10 most cited by the general literature were also compared to assess the extent to which the research in the dataset is representative of the broader research landscape.

Data Comparability and Analysis

For the representativeness of the dataset in terms of content, this was done through the analysis of the evaluation data presented within the publications. Specifically, all throughout the study, the content of the publication were compared with the general literature. In general, the comparison was done in order to clarify and perform the following:

1. **To assess the quality of evidence**: The 3ie DEP focuses on the general evaluations of development interventions using several techniques (impact evaluations, systematic reviews and evidence gap maps), which may provide high-quality evidence of what works in development. By comparing the findings from the portal to the general literature, it may be possible to assess whether the findings are consistent with other research on the same topic.

2. **To identify gaps in the evidence**: By comparing the evidence from the DEP portal to the general literature, it is possible to identify gaps in the evidence base. This may help to identify areas where more research is needed to generate a robust evidence base for policy, practice and further research.

3. **To validate findings**: Comparing the findings from the portal to the general literature can help to validate the findings. If the findings are consistent with other research on the same topic, this adds credibility to the evidence.

4. **To contextualize the findings**: The evidence from the portal may be contextualized by comparing it to the general literature. This can help to understand the broader context of the findings, and to identify factors that may be influencing the results.

To evaluate the data presented within the publications, the key elements, findings and evidence presented in the publications are first identified. Then, a comprehensive literature search on the topic to identify relevant studies and evidence beyond the publications in the 3ie DEP was conducted. In the next step, some or all of these steps were performed, depending on the nature of the considered dataset content:

1. Compared the findings and evidence presented in the publications with the general literature.

2. Validated the findings by comparing them to the general literature. If the findings are consistent with other research on the same topic, this adds credibility to the evidence.

3. Identified gaps in the evidence base by comparing the evidence from the DEP portal to the general literature. This helped to identify areas where more research is needed to generate a robust evidence base for policy, practice and further research.

4. Contextualized the findings by comparing the evidence from the portal to the general literature. This can help to understand the broader context of the findings, and to identify factors that may be influencing the results.

Seeing as the design of the evaluations featured in the papers of this dataset was an essential element to the study's analysis and finding, a separate chapter, Chapter 4, was dedicated to the analysis of the different design features.

Nomenclature

One thing to note is that throughout the analyses, the findings from the dataset and those from the literature were clearly delineated by headings clarifying whether the succeeding revolved around the dataset or the general literature. In certain instances, some clarification was needed for terms or concepts included within the dataset publication, and so external literature was consulted to fill in the gaps. In each one of those cases, it was explicitly denoted that what follows is extracted from sources outside of the dataset.

Furthermore, for the sake of coherence in what follows in terms of nomenclature, a trial is indicated by the in-text citation of the study that centered around it. In the case of trials having multiple associated publications, then the trials themselves were indicated by their trial name to differentiate between them. The details of the multi-publications trials can be found in Appendix A. However, whenever a particular study of one of the multi-publications trials was to be referred to, then the in-text citation of the study and the name of the trial are cited.

3
Results

This chapter details the results drawn from the data mining procedure as well as the bibliometric and network analyses respectively.

3.1 SEARCH RESULTS

In what follows, the results of the sets of publications obtained, one set pertaining to international publications covering areas globally and the other specific to SSA, are reported. The chapter starts with illustrating the results of the publications search process, followed by the initial broad categorization of the fields of the subjects of the studies in both sets.

The initial keyword search strategy, detailed in Chapter 2, produced 98 hits with 66 relating to SSA. The titles and the abstracts were then screened and non-relevant studies were discarded. The final dataset comprised 78 publications of which 48 revolved around SSA. For a better representation of the data mining process and the formation of the dataset, Figure 1 illustrates the different steps taken.

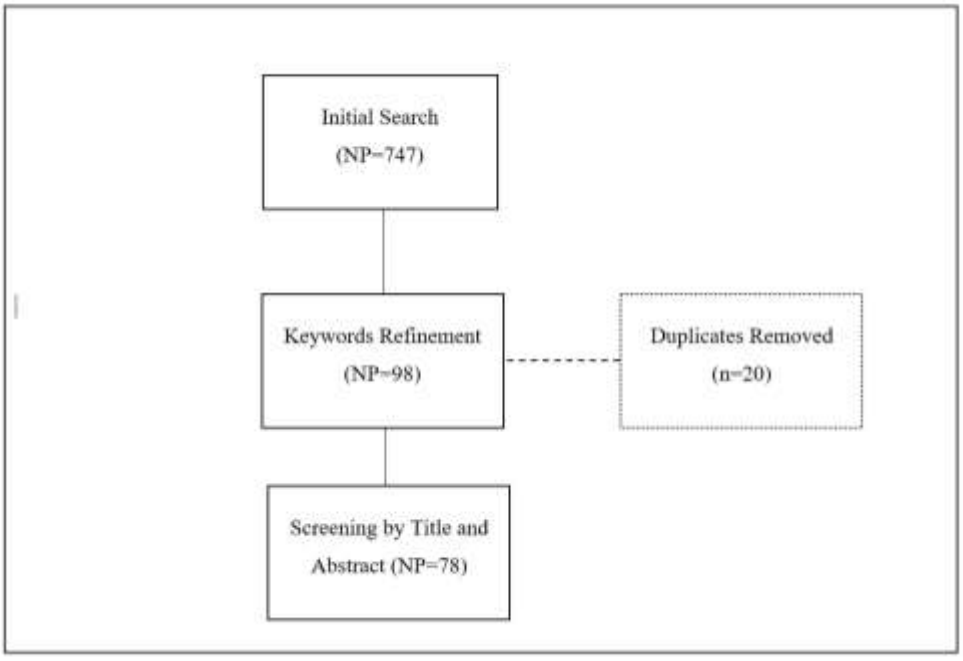

Figure 1: Schematics of Dataset Formation Procedure

While browsing through the studies, two main categories of sectors were identified during the initial broad categorization of the content of the publications in both dataset, and they were related to water and sanitation interventions and agricultural development.

A. TEMPORAL DISTRIBUTION OF SSA AND INTERNATIONAL SECTOR DATA

In order to put the trends observed in SSA with respect to the worldly contributions, the graph in Figure 2 was produced to illustrate the yearly distribution of agricultural-related interventions and those pertaining to the water and sanitation sector, in SSA and the world.

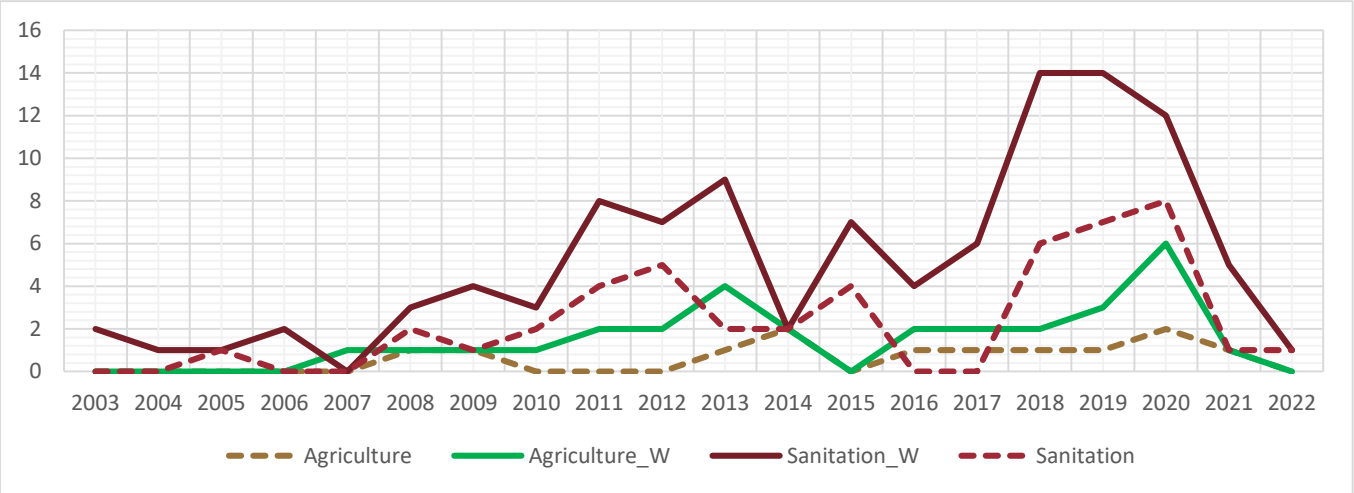

Figure 2: Graph of the General Sector Distribution in SSA & World per Year

In general, both the publications related to SSA and the global one have similar trends to each other for both the agricultural and the WASH sectors, as agricultural productivity seems a less popular sector in both the SSA dataset and the global one. Furthermore, the publication rate is shown in both cases to pick up from the year 2008 and onwards.

B. GEOGRAPHIC DISTRIBUTION OF SSA SECTOR DATA

Zooming onto SSA, the distribution across the countries in SSA of the publications relating to the agricultural and the WASH sectors is displayed in Figure 3.

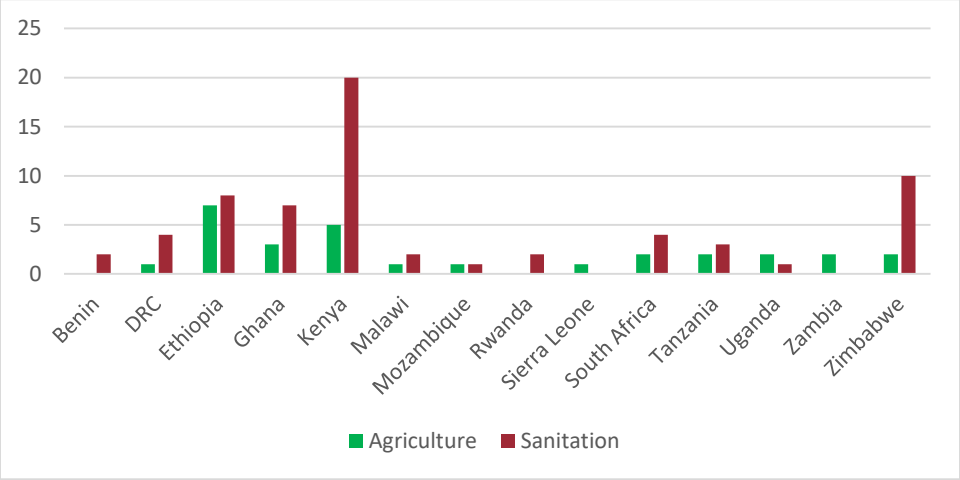

Figure 3: General Sector Distribution/ SSA Country

Of the 46 resultant studies, only 16 revolved around irrigation and agricultural productivity and the rest on WASH interventions. It was decided accordingly to eliminate agriculture studies in order to narrow down the focus of this study to one field. At this point, the set of 30 publications was considered as finalized.

3.2 RESULTS OF THE BIBLIOMETRIC ANALYSIS

In what follows, the results of the bibliometric analysis are detailed. The temporal evolution of the number of dataset publications is first illustrated, followed by the bibliometric findings of each of the three constituents considered in this study: dataset as a whole, journals and authors.

3.2.1 TEMPORAL DEVELOPMENT OF THE DATASET

As can be noted from the graph in Figure 4, the growth trends of the number of featured studies do not exhibit a singular uniform pattern, as they showed fluctuation tendencies across the years. One thing to note is that the oldest study dates back to 2005 by (Crump et al., 2005). Furthermore, two stagnation periods can be identified between 2005 to 2007 and between 2015 to 2017. Beyond those two periods, almost steady growth trends are exhibited.

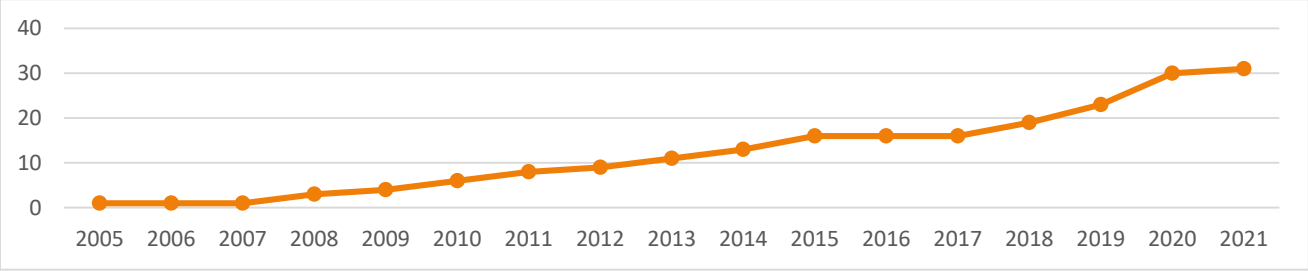

Figure 4: Graph Depicting the Yearly Cumulative Number of SSA Publications

3.2.2 PUBLISHING JOURNALS

A. COUNTRY OF ORIGIN OF THE JOURNALS

The country of origin of the journals, the data is showcased below. The two most productive countries are the UK accounting for 34% of the journals and the United States of America (USA) 31%. Journals from developing countries are sorely lacking, as Ethiopia, Ghana and Nigeria are the only publishing countries, amounting for 3% of the journals each.

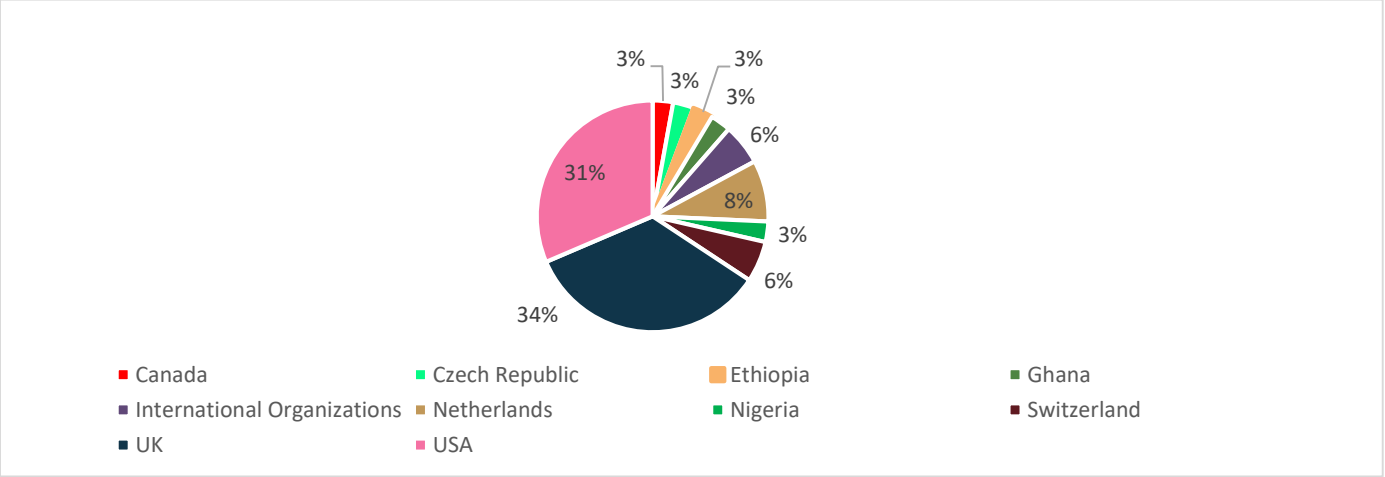

Figure 5: Chart Depicting the Percentage of Publications/ Journal Country

B. FIELD DOMAINS OF THE JOURNALS

Furthermore, the main focus of the journals featured in the dataset and the percentage of articles they published are represented in Figure 6.

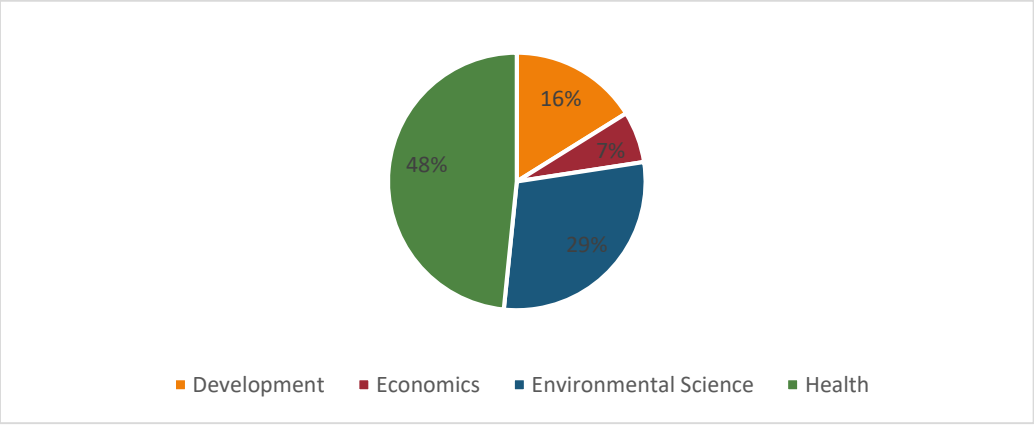

Figure 6: Chart Depicting the Number of Publications/ Journal Focus

All in all, approximately half of the journals (48%) in which the papers of the dataset were published are related to tropical medicine and in particular epidemiology, infections and diseases, health and medical research. While a third of the journals are related to environmental science and rest distributed between development-focused journals (16%) and a minority in economics journals.

C. PRODUCTIVITY AND PERFORMANCE OF THE JOURNALS

The graph in Figure 7 illustrates the number of occurrence frequency and the JCI of the journals featured in the SSA dataset. The only study published in 2022 is too recent and hence was not included in the analysis.

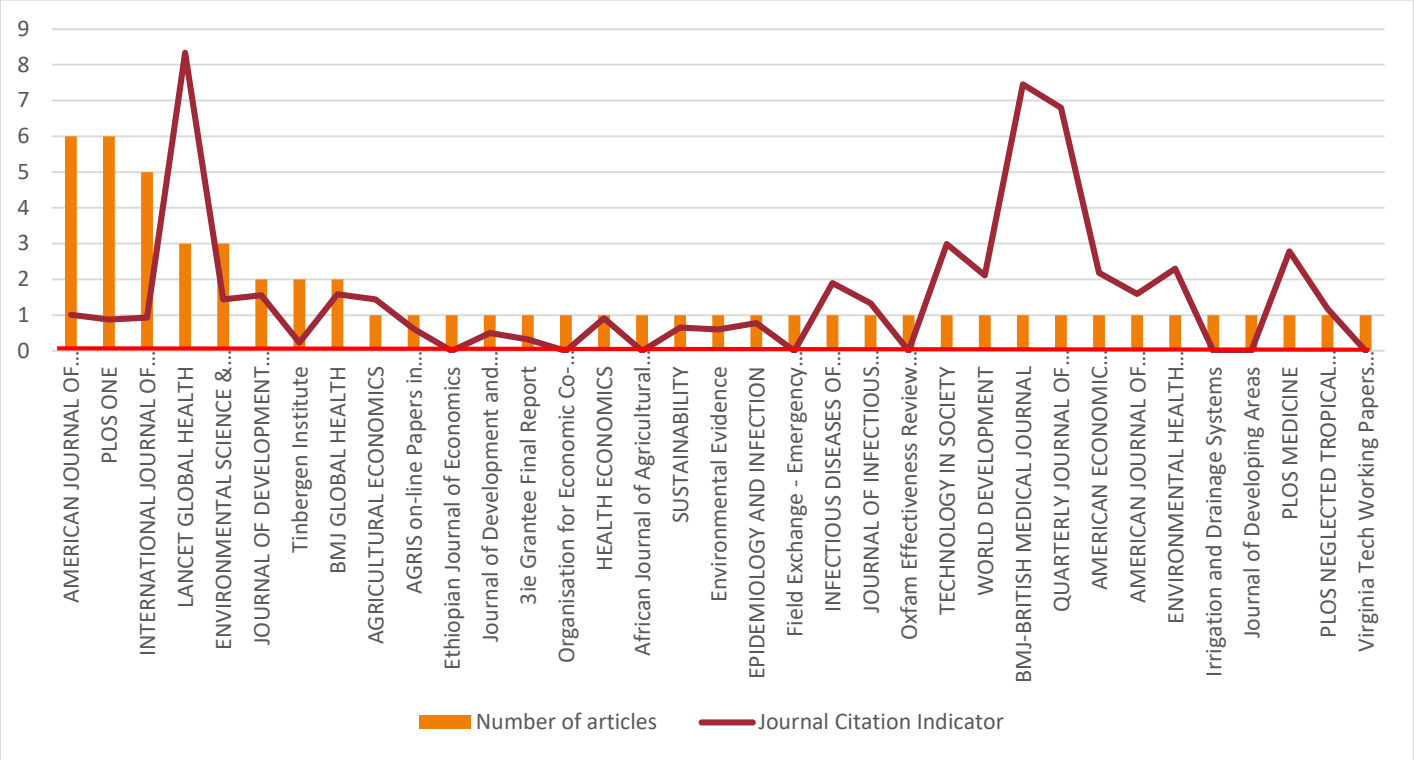

Figure 7: Graph Depicting the Number of Journal Occurrences/ JCI

It is worthy to note that the only two journals that originate from developing countries, which are Ethiopian Journal of Economics (Ethiopia) and Journal of Development and Agricultural Economics (Ghana) have no journal rankings or any information about their global standing.

Concerning the distribution of publications across the journals, 30% of the articles are aggregated between three journals: America Journal of Tropical Medicine and Hygiene (NP = 6; JCI = 1.01), PLOS ONE (NP = 6; JCI = 0.88) and International Journal of Environmental Research and Public Health (NP = 5; JCI = 0.93). The first two journals are from the USA and the last from Switzerland. Their JCI's indicate that these journals published papers that obtained the world citation average in their subject categories. All in all, the number of journals having JCI above than 1 are

almost equal to those with a JCI below 1. As for the publications, 47% were published in journals having a JCI equal to or above 1 and only 16% above 2.

Most Productive Journals

Zooming in on the three most productive journals, the yearly number of publications issued by each one is showcased in Figure 8.

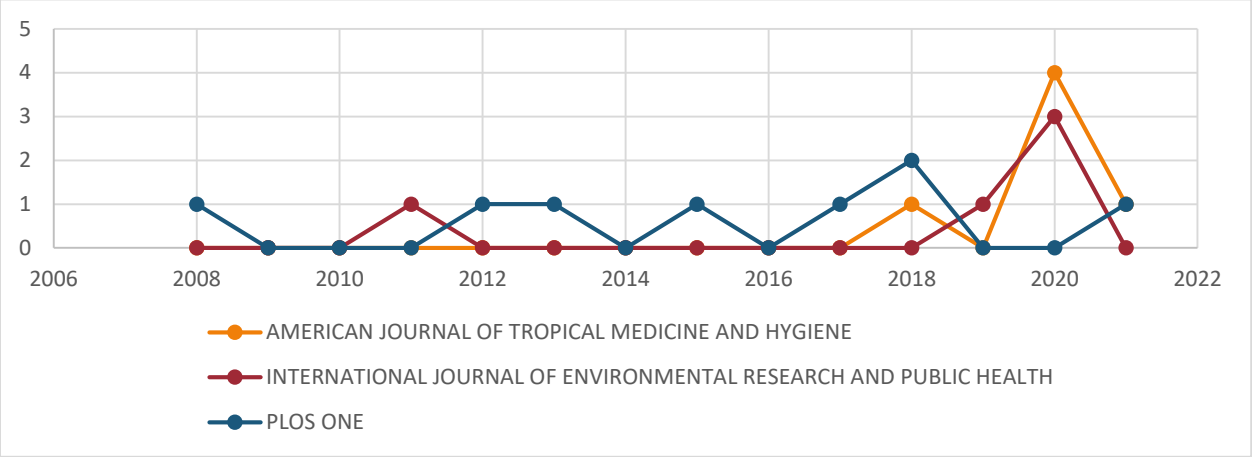

Figure 8: Graph Depicting the Yearly Number of Publications of Top 3 Most Productive Journals

The publications from PLOS ONE journal (NP = 6; JCI = 0.88, USA) were the most spread out temporally and geographically and was the only journal to have published before 2011. It included two articles published in 2015 and 2021 revolving around the same project in Zimbabwe, and common authors that published two articles around different countries and projects. This suggests that the journal may have a larger pool of authors and readers and may have a focus on interdisciplinary and collaborative research

On the other hand, the American Journal of Tropical Medicine and Hygiene (NP = 6; JCI = 1.01, USA) became only active as of 2018. It focused almost entirely on Kenya and includes only one paper on South Africa and Zimbabwe in 2020 and two articles belonging to the same project in 2020, which explains its relatively high number of publications in 2020. The journal may be a newer player in the field, with a specific area of research focus.

Lastly, and although the International Journal of Environmental Research and Public Health (NP = 5; JCI = 0.93, Switzerland) started to be properly active in 2019, it includes diverse papers about different trials in different countries with no recurring authors. This could imply that the journal may be open to a broad range of topics and may have a diverse pool of authors.

3.3 PUBLISHING AUTHORS

A. COUNTRY OF ORIGIN OF THE AUTHORS

The productivity and geographic diversity of the affiliations of the researchers who authored the papers in the dataset are depicted in the graphs in Figures 9 and 10 respectively:

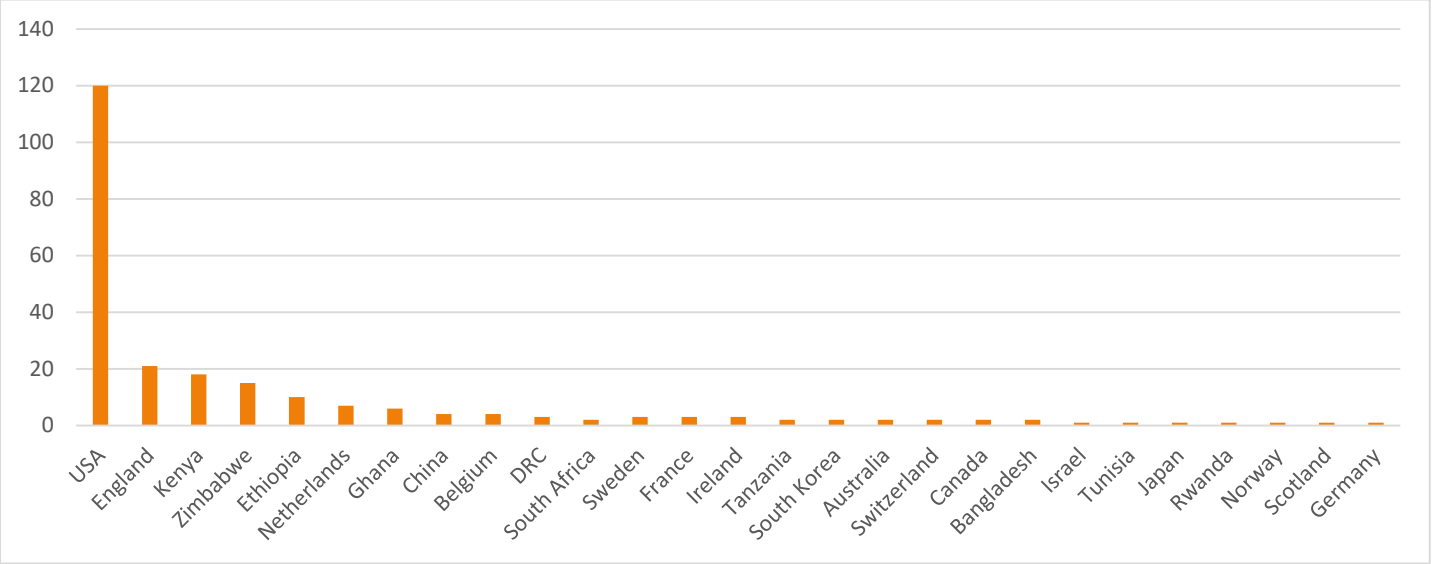

Figure 9: Graph Depicting the Total Number of Featured Organizations/ Country

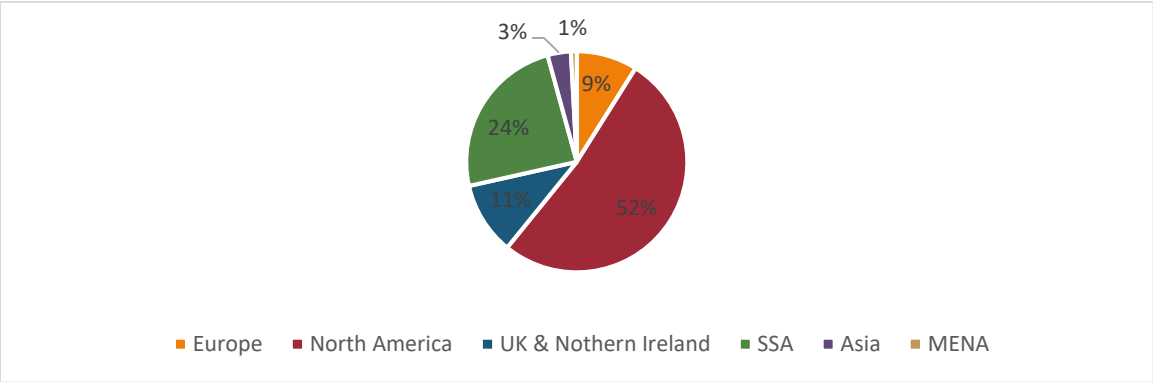

Figure 10: Chart Depicting the Geographic Diversity of Author Affiliations

In total, 243 author organizations were identified, 50% of which reside in the USA and 24% in SSA, 11% in the UK and Northern Ireland, and 9% in Europe. As expected, organizations in Asia and the MENA region have poor publication numbers within the dataset.

In general, only 24% of the organizations are from SSA and they are mainly aggregated between Kenya (7%), Zimbabwe (6%) and Ethiopia (4%). However, as these three countries have the highest number of publications/ country, it follows that they would have the most contributing affiliations amongst those from developing countries.

B. PRODUCTIVITY AND PERFORMANCE OF THE JOURNALS

The number of articles authored per person and the irrespective H-index are illustrated in Figure 11. Due to the high number of authors, the top 20 most productive authors are featured in the graph. One thing to note is that the H-indices values were retrieved on 04/01/2023.

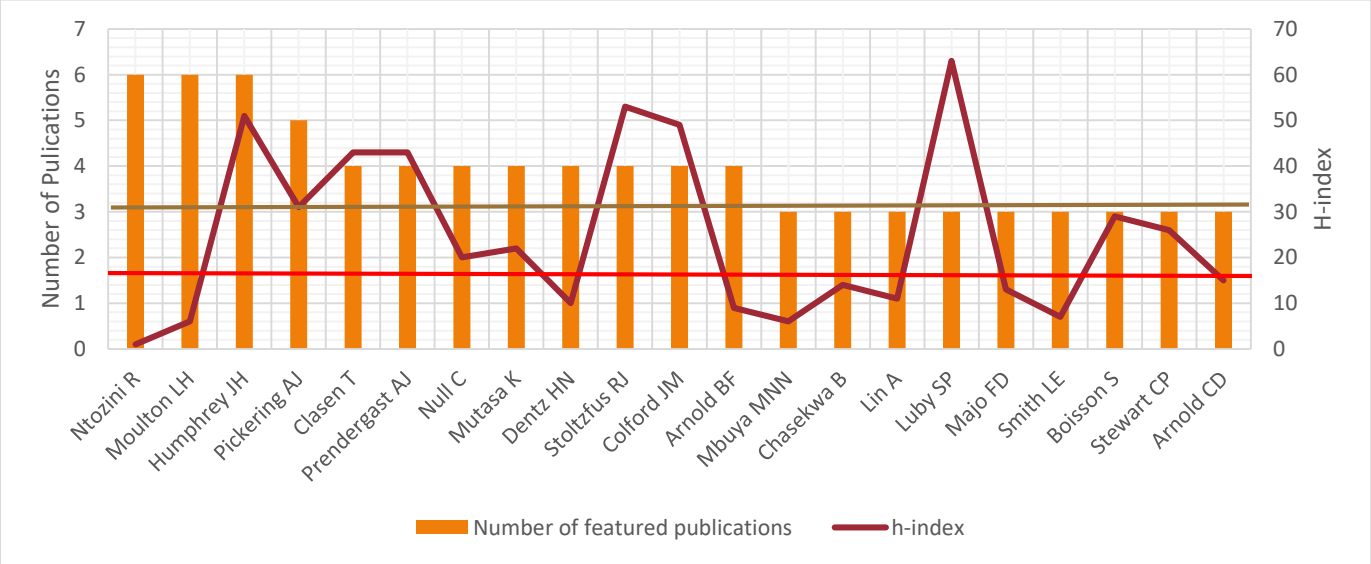

Figure 11: Graph Depicting the Publications Number of the 20 Most Productive Authors and H-index

Firstly, 17 out the top 20 most productive authors in the dataset belong the research teams of the two trials that have the highest number of studies written on them, the SHINE trial with six publications and the WASH Benefits Kenya trial with 4. The details of these trials are featured in Appendix A.

Secondly, and as is illustrated in Figure11, 5 authors have H-indices between than 20 and 40, and 6 authors with H-indices greater than 40. Furthermore, only one author, Luby SP has an H-index greater than 60.

The top four most productive authors are: Humphrey JH (NP = 6, H-index = 51), Moulton LH (NP = 6, P-index = 6), Ntozini R (NP = 29, H-index = 22) and Pickering AJ (NP = 5, H-index = 31). As can be observed, Humphrey JH and Ntozini R had “great” standings while Pickering AJ had a “good” one. Furthermore, Humhrey, Ntozini and Moulton

are all researchers on the SHINE trial team, while Pickering is a recurrent author in the studies related to the WASH Benefits Kenya trial.

As for the field of interest of the four most productive authors, Ntozini, Humphrey, Moulton and Pickering are all concerned with subjects related to tropical medicine, immunology and environmental health. It is worthy to mention that Ntozini, Moulton and Humphrey co-authored all six publications that revolve solely on the SHINE trial in Zimbabwe initiated in 2015. The three authors are in fact part of the research team of the trial. Accordingly, they are affiliated with the Zvitambo Institute for Maternal and Child Health Research in Zimbabwe, while Humphrey and Moulton are also affiliated with the Johns Hopkins University and research centers. As for Pickering, her affiliation is with the University of California and engages in projects mostly in Kenya that revolve around WASH interventions and child health.

3.3 RESULTS OF THE NETWORK ANALYSIS

In the succeeding, the results of the citation and keyword analyses are reported respectively.

3.1 CITATION ANALYSIS

A. CITATIION NETWORK CLUSTERS

The citation network visualization presented by VOSViewer included five clusters in total. One thing to note is that the resolution value used to aggregate the clusters was 2.75.

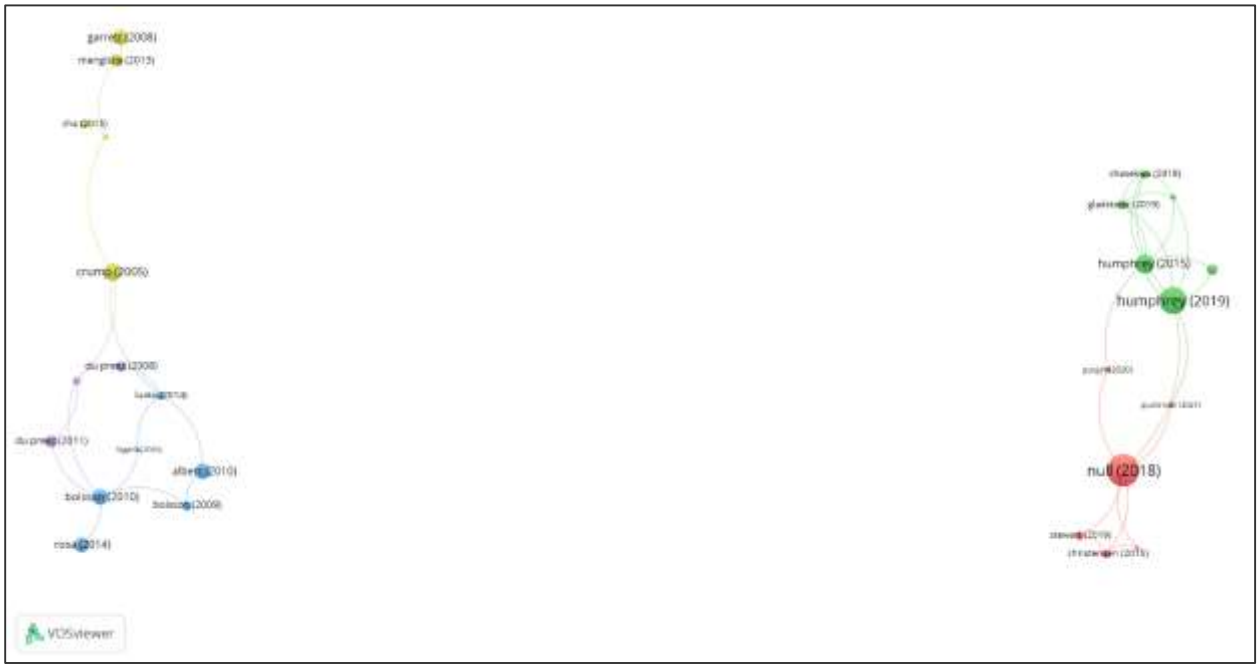

Figure 12: Citation Network of the Dataset Papers

As is evident from the map in Figure 12, two chief separate clusters were formed in the citation network, each encompassing two or three clusters. The three largest Clusters (I), (II) and (III) in red, green and blue respectively with 6 items, Cluster (IV) in yellow with 5 papers and Cluster (V) in purple with 3. Upon closer inspection, each cluster was found to be comprised of papers that were related to each other by common results, intervention investigated or having been performed by the research team. The characteristics of the clusters are detailed below.

Characteristics of the Citation Network Clusters

Cluster (I) and (II):

The papers in the first large grouping of articles, including Clusters (I) and (II) are all comprised of papers pertaining to large-scale WASH trials. Cluster (I) includes 5 publications revolving around a large-scale trial in Kenya called WASH Benefits Kenya, the details of which are detailed in Appendix A, revolving around the effects of WASH interventions coupled with nutritional supplementation. All of the studies are done by the same research group with some common authors found across the different studies. In that cluster, (Null, 2018) is configured as a central publication with considerable connectivity and hence citation connections around it. That paper was also highly cited in external literature, an observation that can be made by its large node size relative to the others around it. Upon closer inspection of the study topic, it was found to revolve around the effects of the trial on diarrhea and stunting, while the 3 of the other connected papers are related to improvements in other health outcomes and one, (Christensen, 2015) was the study related to the pilot trial.

The same observations apply to the Cluster (II) as all of the papers belonging to it revolve around the SHINE trial, which is also detailed in Appendix A, which is why it was found that many of the authors are common among the papers. Furthermore, same as (Null, 2018) in Cluster (I), (Humphrey, 2019) was also a considerably cited paper externally that is well connected in the cluster, and it also revolved around diarrhea prevalence. As for the rest, 4 of the papers revolved around other health outcomes such as stunting, neurodevelopment or diarrhea-inducing gastrointestinal illnesses, while (Humphrey, 2015) was the preliminary publication describing the method, rational and preliminary data of the intervention.

Lastly, and the only publication in the two clusters that was not pertaining to SHINE or WASH Benefits Kenya was (Quattrochi, 2021). The representative node is located centrally between the two clusters, and upon closer investigation, the study was found to have revolved around a large scale trial in the DRC. The intervention approach that the study adopted was one that was different than the other two trials, as it was a community based project that empowered communities to come up with solutions to their self-identified WASH-related issues. The study was found to have cited studies from the two other trials as a recognition of the common characteristic of being large scale trials.

Cluster (III):

Group (III) publications are all studies that included filters in their evaluations. Four of the studies, (Boisson, 2010; Rosa, 2014; Fagerli, 2020; Boisson, 2009) all revolve around the same brand of filters, the LifeStraw hollow-fiber filter, with (Boisson, 2009) investigating the portable version of the filter. The last two studies, (Albert, 2010; Luoto, 2014) were around the same trial and had several researchers in common; they investigated three types of household POU water treatments, one of which was ceramic household filters. Within this cluster, (Boisson, 2010)

exhibits the highest degree of connectivity to other papers, it was found to have been cited for its unique approach of conducting a blinded trial, as opposed to the rest of the studies that were unblinded. In particular, (Boisson, 2010) was cited mostly for justifying the reason the other trials went with an unblinded approach, since (Boisson, 2010) reported to have faced considerable challenges in the blinding process and to eventually having failed in doing so.

Cluster (IV):

Four out of the 5 publications in Cluster (IV) were trials that investigated POU water treatments and their effectiveness in reducing diarrhea. They all included chlorination except for (Crump, 2005) who used flocculent disinfectant. (Cha, 2015) on the other hand focused on water supply and sanitation for the reduction of diarrhea, and indeed the node belonging to it could be noticed to branch out from the man chain of publications formed. Furthermore, all of the studies in this cluster obtained positive results for the reduction in diarrhea.

While exploring the reasons for the citations in this cluster, it was made clear that all of the studies, except for (cha, 2015), built upon (Crump, 2005)'s use of flocculent-disinfectant to reduce diarrhea and the study's positive results to motivate their choice of using the use of another household POU water treatment, chlorination, to reduce diarrhea. This could also be verified from the map, as (Crump, 2005) was located at the tip of the chain.

Cluster (V):

Cluster (V) includes three studies, two of which, (du Preez, 2008; du Preez, 2011), were performed by the same research team investigating diarrhea reduction using ceramic filters and solar disinfection respectively. (Stauber, 2012) on the other hand studied the effects of biosand filters on diarrhea reduction. The first studies were cited by (Stauber, 2012) in order to validate the positive results obtained.

Finally, one general trend observed is that apart from the aforementioned citation reasons, studies in this dataset validate or question their results based on the findings of the other studies.

B. INTERNALLY CITED ARTICLES

Twenty-five out of the 30 publications in the dataset cited each other at least once. However, their local standing within the research community in the dataset may not be concurrent with their global one in the external general literature. As such, and to get a clearer idea of the degree of influence of the data used in this study, the global citations and internal links of the top 10 most cited articles globally and locally are presented in Tables 1 and 2 respectively.

Table 1: Most Cited Publications Globally

Articles	Global Citations	Network Links	Journal	JCI
Null et al. (2018)	305	8	Lancet Global Health	8.34
Humphrey et al. (2019)	220	8	Lancet Global Health	8.34

Pickering et al. (2015)	210	5	Lancet Global Health	8.34
Kremer et al. (2011)	129	2	The Quarterly Journal of Economics	6.8
Humphrey et al. (2015)	103	6	Clinical Infectious Diseases	1.9
Crump et al. (2005)	89	3	BMJ	7.45
Albert et al. (2010)	76	2	Environmental Science and Technology Journal	1.44
Boisson et al. (2010)	73	5	PLOS One	2.78
Garett et al. (2008)	68	1	Epidemiology and Infection	0.78
Rosa et al. (2014)	51	1	PLOS One	2.78

Table 2: Most Cited Publications Within the Dataset

Articles	Global Citations	Network Links	Journal	JCI
Humphrey et al. (2019)	220	8	Lancet Global Health	8.34
Null et al. (2018)	305	8	Lancet Global Health	8.34
Humphrey et al. (2015)	103	6	Clinical Infectious Diseases	1.9
Boisson et al. (2010)	73	5	PLOS One	2.78
Chandna et al. (2020)	10	4	BMJ	7.45
Gladstone et al. (2019)	24	4	PLOS One	2.78
Chasekwa et al. (2018)	22	4	PLOS One	2.78

Christensen et al. (2015)	21	4	American Journal of Tropical Medicine and Hygiene	1.01
Luoto et al. (2014)	21	4	Journal of Development Economics	1.56
Stauber et al. (2012)	22	4	International Journal of Environmental Research and Public Health	4.61

Firstly, when looking at the JCI's of the journals in Table 1, it can be noted that all the papers were featured in journals with a good standing (JCI >1). Furthermore, the two most cited articles in the dataset were featured in a journal with an "excellent" academic standing (JCI >5).

When zooming on Table 2, the global citations values and the internal network links of the publications in the dataset are adequately correspondent to each other. By comparing the two tables, the top five most cited articles within the dataset network are featured in the top 10 most globally cited publications amongst those in the dataset. These observations could therefore indicate that the set of data in this study is likely a representative sample of the research field of the evaluations of aid interventions in the water sector.

C. INTERNALLY NON-CITED ARTICLES

Internally non-cited articles could either be recent, so they would not have had enough time to be recognized and to assemble citations. They could also be not prominent in their respective fields or simply not related to the main areas of focus of the dataset. They could also represent a literature gap within the dataset (Smith et al., 2019).

As for the documents that were not cited by any other publications within the dataset, their information is depicted in Table 3:

Table 3: Internally Non-Cited Publications

Publication	Global Citations	Journal	JCI
Fisher et al. (2020)	1	PLOS One	2.78
Morris et al. (2018)	9	American journal of tropical medicine and hygiene	1.01
Gunther et al. (2013)	37	Health Economics	0.91

Patel et al. (2012)	35	American journal of tropical medicine and hygiene	1.01

The characteristics of these studies were reviewed to detect any disparities with the other publications such as publishers, funders or affiliations but nothing was found. Next, their content was reviewed and indeed the approaches or interventions included in three of them were different.

Firstly, (Morris et al., 2018) was differentiated from the rest of the internally cited studies in the dataset, as they investigated the effect of colloidal silver ceramic filters on *Cryptosporidiosis* prevalence in the water. The choice of treatment was different than the rest of the ones in the dataset as the filters were colloidal-silver coated. The paper was also the only one to have studied *Cryptosporidium* contamination reduction.

Likewise, Gunther (2013) was the only study that focused solely on improved water storage and transportation, investigating their effects on water quality improvement and household health. Finally, Patel (2012) was the only study that was done in a school setting as opposed to the common household settings in the dataset. The study was also one of the few studies in the dataset to have investigated respiratory infections, and it was the only one that explored the effects of improved water supply in schools coupled with the implementation of a hygiene curriculum on the decrease in the risk of respiratory infections.

On the other hand, three of studies, (Morris, 2018; Dockx, 2019; Fisher, 2020) were published in 2018 or later on, which might explain their low global citation count. This may also explain the reason behind them not being cited by any study in the dataset, despite them being published in journals with relatively well academic standings.

3.2 THEMATIC ANALYSIS

Through the thematic network generated and visualized by VOSViewer, a total of three clusters was obtained from the top 50 most co-occurring keywords in the dataset. The resolution number used to obtain the clusters was 1.75. The resulting map is displayed in Figure 13:

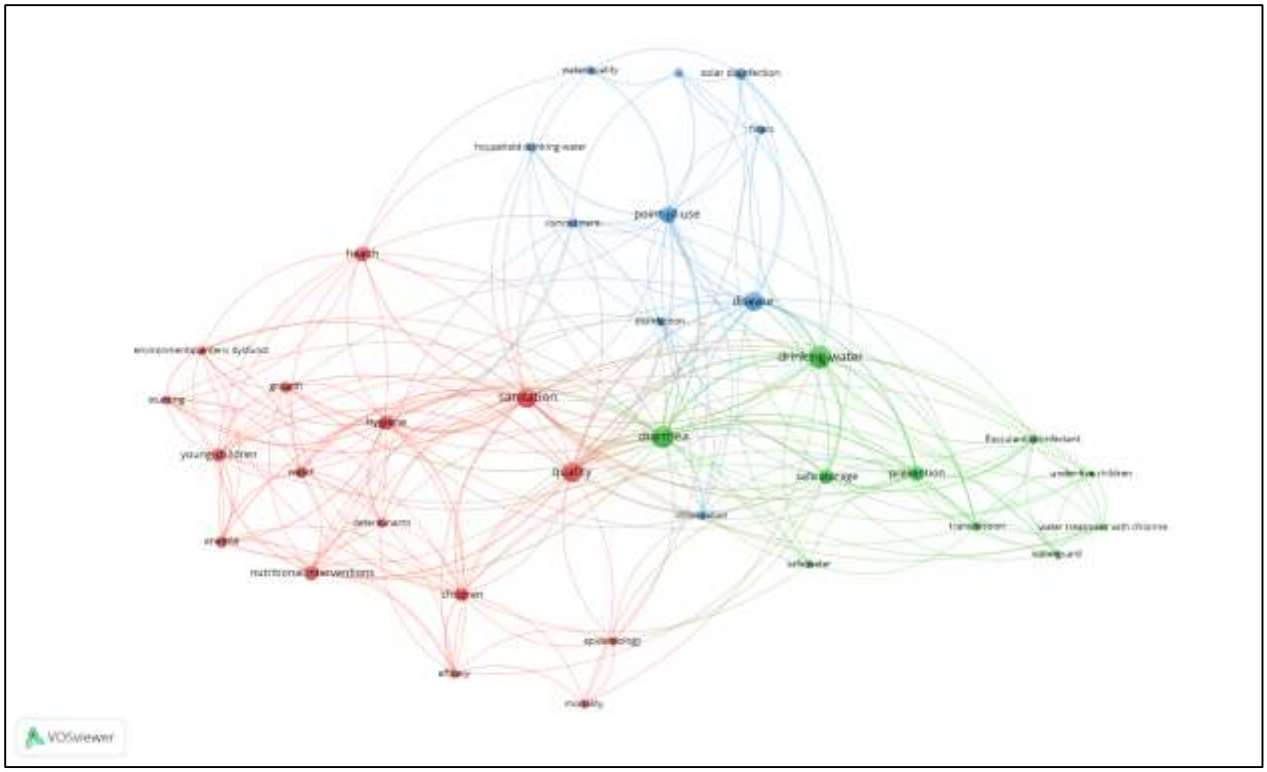

Figure 13: Keyword Network Map

Furthermore, the ten words that most frequently appeared with other terms are shown in Table 4 that also indicates the total link strength of the terms and the number of their clusters:

Table 4: Top 10 Most Recurrent Keywords

Keyword	Total link strength	Cluster
Diarrhea	72	2
Drinking-water	67	2
Sanitation	55	1
Quality	47	1
Disease	45	3

Hygiene	37	1
Prevention	37	2
Safe storage	37	2
Point-of-use	34	3
Children	26	1

Firstly, the fact that the top keywords include "Diarrhea," "Drinking-water," "Sanitation," and "Hygiene" suggests that the researchers were interested in understanding the impact of WASH interventions on public health outcomes.

Characteristics of the Thematic Analysis Clusters

In total, three clusters were obtained that indicate the following themes within the dataset:

Cluster (I): Cluster I in red includes keywords such as anemia, children, stunting, environmental enteric dysfunction, sanitation and nutritional interventions. These words are particularly indicative of the two large-scale trials, SHINE and WASH Benefits Kenya, that share the same approach for health outcome improvement, particularly via sanitation and nutritional supplementation. These trials have multiple studies revolving around them, and these studies form a subset of literature further demonstrated by the clustering in the citation analysis.

Cluster (II): Cluster II in bleu is centered around the keyword point-of-use and disease, and includes terms such as household water-drinking, solar disinfection, filters, chlorination, disinfection and commitment.

The inclusion of keywords such as “point of use”, "household water-drinking", "solar disinfection", "filters", "chlorination", and "disinfection" all suggest a focus on exploring the effectiveness of various household water treatment options in improving household drinking water quality.

The keyword "commitment" stands out as it is not directly related to household water treatment options but is recurrent, closer to the keyword "point-of-use" and well connected. It is possible that the term "commitment" may refer to the recognition of the need for households to make a sustained effort to implement and maintain household water treatment options, and so user commitment could be a major theme in the dataset. However, the word “commitment” may also refer to the recognition for the need for commitment from government agencies, NGO’s, and other stakeholders to support the implementation and maintenance of household water treatment options.

Cluster (III): Cluster III in green is centered around the following three words in order of co-occurrence: drinking-water, safe storage and prevention. The cluster also includes the following indicative words: diarrhea, flocculent disinfectant, water treatment with chlorine, transmission, and Waterguard. The last term is a brand of chlorine common in SSA.

While some of the indicative keywords in this cluster may relate to household water treatments, and so the dominant theme of Cluster (II), two observations can be made: the cluster only includes "chlorination/Waterguard" and "flocculent disinfectant" as water treatment hardware, and one of the main keywords is "water storage". This suggests that the focus of this cluster may be on evaluating the effectiveness of safe water storage practices and/or addressing issues related to water contamination, and their effect on diarrhea.

The inclusion of the keyword "prevention" further supports the idea that the focus of this cluster is on preventing the transmission of waterborne diseases through safe water storage practices.

3.4 BIBLIOMETRIC REPRESENTATIVENESS OF THE DATASET

In what follows the procedure and results drawn from the analysis of the comparability and subsequent representativeness of the dataset is detailed.

A. INITIAL RESULTS

For evaluating the degree to which the dataset can be deemed as representative of the general literature, and using the keywords indicted in Chapter 2, the first search on WoS produced a total of 38,160 publications.

The main characteristics of the results are displayed in Table 5, as extracted from WoS categorization:

Table 5: Characteristics of generalized search (no study design specification)

TOPIC CATEGORIES	AUTHORS	AFFILIATIONS	JOURNAL COUNTRY	JOURNAL
ENVIRONMENTAL SCIENCES	Stephan, LP	University of London	USA	Science of the Total Environment
WATER RESOURCES	Janet, S	London School of Hygiene Tropical Medicine	China	Sustainability
PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH	Clasen, T	Ministry of Agricultural Rural Affairs	England	PLOS ONE
AGRONOMY	Hayes, JH	Chinese Academy of Sciences	Australia	American Journal of Tropical Medicine and Hygiene
GIOSCIENCES MULTIDISCIPLINARY	Todd, J	CGIAR	Germany	International Journal of Tropical Medicine and Hygiene

TROPICAL MEDICINE	Gassman, P	University of California System	Canada	Agriculture Water Management
GREEN SUSTAINABLE SCIENCE TECHNOLOGY	Mabey, D	Chinese Academy of Agricultural Science	Spain	Journal of Environmental Management
ENVIRONMENTAL ENGINEERING	Unicomb, LE	Wageningen University Research	France	Environmental Science and Pollution Research
METEOROLOGY	Colford, JM	Johns Hopkins University	Netherlands	BMJ
FOOD SCIENCE TECHNOLOGY	Simon, C	Centre National de Recherche Scientifique	Italy	Malaria Journal

As can be deducted from Table 1, only two topic categories are related primarily to Health: Public environmental occupational health and Tropical medicine. The rest of the categories are either related to agriculture and water access, water quality, environmental sciences or food security.

B. RESULTS OF THE REFINED SEARCH STRATEGY

In order to further restrict the search, and since all of the articles in the dataset were found to be revolving around randomized controlled trial, the following terms were added in WoS search engine as keywords to be present in the title, abstract or article keywords: (randomized controlled OR randomized-controlled). With this inclusion, the search narrowed down to 232 hits, the top characteristics of which are illustrated in Table 6.

CHARACTERISTICS OF THE WoS RESULTS

As can be observed from Table 2, the addition of the trial property restriction has caused the most recurrent topics of the publications to shift to health, medicine and nutrition. That was to be expected as the implementation of the specified type of clinical trials is most common in health and medicine fields (Heneghan et al., 2017).

As for agriculturally and ecologically-focused publications, they are only 29 out of the 232 hits, and are mostly about small-scale irrigation infrastructure. The countries of focus of these studies are mostly Ethiopia, Ghana, Kenya respectively. As for the publishing journal countries, they are respectively as follows: UK, Germany, Netherlands, USA, Switzerland. It could be noted that the areas of focus of the global dataset and the internal dataset are similar.

Table 6: Characteristics of design-specific studies search

TOPIC CATEGORIES	AUTHORS	AFFILIATIONS	JORNAL COUNTRY	JOURNAL
PUBLIC ENVIRONMENTAL	Luby, SP	University of London	England	American Journal of

OCCUPATIONAL HEALTH				Tropical Medicine and Hygiene
TROPICAL MEDICINE	Colford, JM	London School of Hygiene Tropical Medicine	USA	TRIALS
MEDICINE GENERAL INTERNAL	Clasen, T	University of California System	Australia	Lancet
INFECTIOUS DISEASES	Arnold, B	Johns Hopkins University	Switzerland	PLOS ONE
NUTRITION DIETETICS	Unicomb, LE	University College London	Kenya	BMJ Open
ENVIRONMENTAL SCIENCES	Stewart, CP	Stanford University	Uganda	BMJ Public Health
PARASITOLOGY	Schmidt, WP	International Center for Diarrheal Disease Research	South Africa	International Journal of Environmental Research and Public Health
MEDICINE RESEARCH EXPERIMENTAL	Rahman, Md M	WHO	Tanzania	Malaria Journal
MULTIDISCIPLINARY SCIENCES	Pickering AJ	Center for Disease & Control Prevention	Netherlands	PLOS ONE Neglected Tropical Health
AGRONOMY	Ram, P	Kenya Medical Research Institute	Gambia	PLOS ONE Medicine

CHARACTERISTICS OF THE DATASET RESULTS

As for the results of the classification of the characteristics of the dataset, they are featured in Table 7.

Table 7: Characteristics of dataset publications

TOPIC CATEGORIES	AUTHORS	AFFILIATIONS	JOUNRAL COUNTRY	JOURNAL	TOP CITED AUTHOR
DIARRHEAL DISEASES	Humphrey, JH	University of California System	USA	American Journal of Tropical Medicine and Hygiene	Clasen, T
NUTRITION DIETETICS	Moulton, LH	Center for Disease Control & Prevention	England	PLOS ONE	Luby, SP

TROPICAL MEDICINE	Ntozini, R	University of California Berkeley	Zimbabwe	Environmental Science and Technology	Arnold, B
EPIDEMIOLOGY	Stoltzfus, RJ	University of London	Kenya	International Journal of Environmental Research and Public Health	Humphrey, HJ
	Mutasa, K	Stanford University	Canada	Lancet Global Health	Quick, RE
	Mbuya, M	London School of Hygiene & Tropical Medicine	Ethiopia	Infectious Disease of Poverty	Pickering, AJ
	Majo, FD	Kenya Medical Research Institute	Ireland	Health Economics	WHO
	Chigumira, A	Johns Hopkins University	South Africa	Epidemiology and Infection	Schmidt, WP
	Tavengwa, NV	Johns Hopkins Bloomberg School of Public Health	Ghana	BMJ Global Health	Hunter, PR
	Jones, AD	University College London	Sweden	Clinical Infectious Diseases	Fewtrell, L

COMPARING THE CLASSIFICATION RESULTS OF THE WOS SET AND THE STUDY DATASET

By comparing Tables 6 and 7, the following observations were made:

- Five out of the ten most publishing journals are common between this study's dataset and the data from WoS;
- Eight out of ten of the most productive affiliations are common between the two sets.
- As for the author countries, four out of ten are common between the sets. Those four countries are: USA, England, Kenya and South Africa;
- Finally, five of the authors that are most cited by articles within the database are featured among the top 10 most productive authors globally. Those authors are: Luby (Times cited= 17,457; H-index= 64, Stanford University), Clasen (Times cited= 7,670; H-index= 44, Emory University), Arnold (Times cited= 3,918; H-index= 34, University of California System), Schmidt (Times cited= 7,059; H-index= 42; University of London) and Pickering (Times cited= 3,515; H-index= 32, University of California System). Finding this particular group of authors may be evident, since they almost always co-author the articles that dataset publications reference. Also, WHO is featured as one of the most cited sources within the dataset, while the organization is also of the most active affiliations globally.

All of the aforementioned suggests that that the publications within the dataset cited authors that may be leaders in their field given their high number of citations and their h-index. Furthermore, the journals in common could mean that the publications included in the dataset are closely related to the overall research and its trends being done in the field.

However, seeing as the inclusion criteria and the study characteristics are considerably narrow, these commonalities could also mean that the dataset is focused on a narrowed down subfield of the general field, in this case, assessments designed as randomized for water-related projects that are specific to a single region, SSA, in which case the journals in common can be specific to that subfield.

This theory though could be contested, since by comparing the Table 1 that did not narrow down the search to a certain type of trials with the tables of the narrowed down search, Table 2, and the table belonging to the articles within the dataset, Table 3, several journals, authors and affiliations are in common. Furthermore, it is also important to consider that the dataset might be biased towards the most popular or well-known journals in the field, and may not include research from less well-known or lesser-cited sources.

All in all, the academic standing of the dataset has been established and its representativeness demonstrated, which means that it was deemed suitable to be the subject of focus of this review.

4

Results of the Content of the Dataset

The following chapter elaborates on the different elements of the trials in the dataset, including the hardware and software of the interventions, the designs of the trial setups, the investigated outcomes and finally the results of the trials.

4.1 INTERVENTION HARDWARE AND SOFTWARE COMPONENTS

Three main types of interventions were featured in the dataset. The most recurrent one was point-of-use (POU) household water treatment that utilized traditional treatment approaches, including chlorination, filtration, flocculation-coagulation and solar disinfection. The second type of interventions was one related to the provision of basic sanitation hardware, while the third is the combination of any of the aforementioned two, including sometimes hygiene and health messaging and/or, infant and young child feeding (IYCF) interventions. All the aforementioned is detailed in the following.

4.1.1 HOUSEHOLD WATER TREATMENT

Four types of POU household water treatments featured in the dataset, and by order of frequency they are: chlorination, filtration, solar disinfection and flocculation-coagulation. The treatments are further detailed in the following:

A. Chlorination: Chlorination was the most common treatments. It was featured in seven of the studies in the dataset, and was recurrent all along the dataset time span and in varied countries, which indicates that it was a consistently well-favored household water treatment option spatially and over time.

The effects of the treatment were studied in isolation in only two studies (Mengistie, 2013; Solomon et al., 2020). In the remaining of the studies, chlorination was featured with other treatments, including water supply and safe storage to reduce re-contamination risk (Garrett et al., 2008). It was also one of three POU water treatments (other two are hollow-fiber filters and flocculation disinfection) provided to households with hygiene messaging to record their preferences and uptake (Albert et al., 2010) of the CARE Kenya trial as well as the effects of two psychosocial messages to evaluate the influence of different messaging techniques on uptake (Luoto et al., 2014), a complementary study of the CARE Kenya. Lastly, chlorination was also one of the multiple interventions offered by the SHINE and Kenya Benefits trials, whereby the treatment was combined with nutritional supplements and

counselling, sanitation and hygiene hardware (soap and latrine construction/rehabilitation respectively) and messaging.

Chlorination in the dataset was always in liquid form (as opposed to tablets) and involved pouring one bottle cap of 1.0% sodium hypochlorite solution into contained water, including water reserved for storage, shaking well the container and waiting at least 30 minutes before intake. All studies studied the effect of providing free chlorine bottles on the reduction of diarrhea in children aged 5 years or less. Furthermore, in all studies, the same benefits of the treatment were repeated, and they included it being effective in improving water quality, inexpensive, easily accessible, simple to implement, fast to treat water and finally having residual protection which would allow it to be used on stored water. Concurrently, the same constraints for uptake were listed in all studies, mainly its residual pungent odor and smell. One study in the dataset, (Solomon et al., 2020), done for 4 months on 405 randomized households in rural eastern Ethiopia, reported a considerable rate of non-compliance (32%) by caregivers to chlorination because of their children not liking the smell or taste of the chlorine in the water and having severely upset stomachs.

B. Filtration: Filtration was the second most recurrent treatment, featured either as ceramic porous filters in five studies (du Preez et al., 2008; Rosa et al., 2014; Morris et al., 2018; CARE Kenya), hollow-fiber filters in three studies (Boisson et al., 2009; Boisson et al., 2010; Fagerli et al., 2020) or biosand filters in one study (Stauber et al., 2012). While all three are types of household water filters, they differ in their design and mechanism for removing contaminants from water.

A biosand filter is typically made up of a plastic or concrete container filled with layers of sand and gravel. Water is poured through the top and flows slowly through the sand and gravel layers, which trap and remove impurities. The filter also contains a biolayer of microorganisms that helps to break down contaminants and further purify the water.

A ceramic porous filter works similarly to a biosand filter, but instead of sand and gravel, it uses a ceramic filter element with tiny pores to remove contaminants from the water as it passes through. One ceramic filter study in the dataset done by (Morris et al., 2018), used a special type of ceramic filter, a colloidal silver ceramic filter. The latter product is a ceramic filter that is infused with silver nanoparticles, which have antimicrobial properties. The silver in the filter is intended to kill or inhibit the growth of bacteria and other microorganisms in the water, adding an extra layer of protection against microbial contamination. In the case of the study, the filter was tested specifically in the removal of *Cryptosporidium*, a chlorine-tolerant protozoan parasite, which was reported by the authors to be the primary cause of moderate-to-severe diarrhea in Kenyan children. More information on this pathogen is featured in Section 4.2 that explores the details of the targeted outcomes of the trials.

In contrast to the two aforementioned filtration systems, a household hollow-fiber filter typically uses a cartridge filled with semi-permeable membranes that remove impurities from the water as it passes through. All three hollow-fiber studies revolve around the same brand of filters, the LifeStraw filter, which is manufactured by a Swiss company called Vestergaard, and distributed in SSA mainly via NGO's partnered with the company. One of the three studies, (Boisson et al., 2009) studied the portable version of the hollow-fiber filter.

C. Solar Disinfection: The third type of POU household treatment was solar disinfection, which involves filling transparent polyethylene terephthalate (PET) bottles with water and leaving them in the sun for solar UV and

increased temperature disinfection. This type of intervention was featured in only one study, (du Preez et al., 2011). Households were provided with two 2-L PET bottles, and were advised to keep the bottles exposed for longer than 6 hours of unobstructed sunlight and to keep storage for a maximum of 48 hours for fear of recontamination. Households were also advised to drink directly from the bottles instead of cup or other containers from fear of recontamination as well.

D. Flocculation-coagulation: Finally, only one study focused solely on flocculation-coagulation, and it happened to be in the oldest study in the dataset, (Crump et al., 2005).

The other time that flocculation-coagulation was featured was in two studies, (Albert et al., 2010) and (Luoto et al., 2014), that belong to the same trial, the CARE Kenya trial. The first one, (Albert et al., 2010), focused solely on user preference between three freely provided products, the other two being chlorination and ceramic filtration. The study was done in two different areas: rural western Kenya and urban Dhaka, Bangladesh. The former was a rural area while the later was an “urban slum”. Despite the incongruities that may arise from the two very different environments, the researchers assessed that there were no major or influencing factors between them, and so were deemed to be rightly comparable, counting in socio-economic determinants and baseline conditions, including the initial state of water and sanitation conditions (for example in terms of contamination level). All households had the chance to try each POU for two months. The free POU were coupled with what the authors termed as “marketing nudges” on usage, which were health-oriented marketing messages inspired by behavioral economics. In both settings, households were subjected to a random assignment to receive varying message "frames" that highlighted the importance of treating water safely. Additionally, half of the households in both settings were randomly selected for a "commitment with reminder" treatment, which entailed a request for a verbal pledge to utilize the safe water product, and visual cues reminding them of this pledge, which they could hang in their homes.

The second study, (Luoto et al., 2014), focused solely on the changes in POU uptake of the three products, through self-reported usage and visual inspection of the products for signs of usage, done three times every two months during household visits to collect data and change treatment product.

4.1.2 DRINKING WATER SUPPLY AND STORAGE

A. Drinking Water Supply: In the dataset, only two trials targeted water supply. The first one was (Cha et al., 2015) that focused on water supply at community level through providing improved drilling and rehabilitating boreholes. The trial also implemented sanitation provision through the construction of latrines in public areas, and evaluated the changes in diarrheal occurrence at households with at least one child under five years. The main focus of the study though was the effects of community water supply.

While the second, (Patel et al., 2012), adopted a unique approach in that the intervention was applied in 42 schools in Kenya whereby a hygiene curriculum was provided along with the construction of new safe water points, the provision of soap for the stations and chlorine bottles for the treatment of the water. The water points themselves were locally produced and consisted of metal stands supporting plastic buckets with a capacity of 60 liters. The buckets were mounted with a lid and a tap attached to them. Student households received biweekly visits by field operators whereby self-reported diarrhea incidence and hygiene practices indicators of the children and their

caregivers were recorded. By doing so, the trial opted to reach out not only to the children at the schools but also their caregivers and other family members through a potential spillover effect.

B. Storage Vessels: As for improved storage interventions, only one study focused on the effectiveness of providing water storage apparatus in the study by (Gunther et al., 2013). Indeed, (Gunther et al., 2013), focused solely on safe water storage through improved household water containers and plastic containers for safe water transport. They evaluated the impact of this hardware on the improvement of water quality through the reduction of *Escherichia Coli* (*E. Coli*) contamination between water from the POS and at household POU, along the reduction of self-reported diarrheal episodes.

The authors chose to conduct the trial in villages with relatively good water source quality. Their choice was justified by noting that it would be impractical to provide improved water vessels in areas where households do not have access to an upgraded water source because improved water containers only maintain the cleanliness of water and cannot purify contaminated water. Secondly, even if it was challenging for policymakers to determine which households have an improved water source, it should be possible to identify villages with access to an enhanced public water source.

4.1.3 SANITATION

There were four sanitation interventions in total in the dataset:

1. SHINE and WASH Benefits Trials: These trials were large-scale studies that evaluated the effectiveness of nutritional supplementation with or without the singular and combined effects of multiple WASH interventions on children health, growth and stunting. The WASH interventions included the installation or rehabilitation of latrines for safe sanitation. The specific details of the trials are elaborated in Appendix A.

The two remaining studies, (Garrett et al., 2008) and the VEA trial introduced hereafter revolved around trials that were particularly distinguished in that the designs of the interventions that they evaluate are two of the most context-oriented and community-based within the dataset.

2. (Garrett et al., 2008): The study implemented a community-wide project among a high-risk rural population in Kenya. The study utilized three types of interventions: hygiene messaging, latrine construction through providing hardware and installation know-how at 40% subsidized cost for households who expressed interest, and rainwater harvesting massaging. The project started off with a two-stage community mobilization approach. It first oriented local government officials, community management committees, and leaders of women's groups to the project to gain their support, and then trained community volunteers to use participatory hygiene and sanitation transformation (PHAST) methodology to promote hygiene, sanitation and rainwater harvesting in community meetings, community management committees, and during home visits.

PHAST was used to engage the community members in identifying and analyzing hygiene and sanitation problems in their community and developing their own solutions. The approach involved a series of participatory activities and tools to engage the community members, such as participatory learning and action (PLA) exercises, community mapping, and participatory hygiene and sanitation assessment (PHSA). These activities aimed to empower the community members to take ownership of the hygiene and sanitation interventions and to promote long-term

sustainability. The study also used the Community-led total sanitation (CLTS) approach, which is a key component of the PHAST approach. CLTS aims to eliminate open defecation by encouraging community members to construct and use their own latrines. The approach relied on community mobilization and triggering to encourage behavior change.

3. (Quattrochi et al., 2021): The study revolved around the 'Healthy Villages & Schools' trial (in French '*Villages et Ecoles Assainis*') (VEA), which is the largest WASH program implemented by UNICEF globally and comprised 90% of total external funding committed to rural WASH in the Democratic Republic of Congo (DRC) from 2005 to 2020.
The intervention employed a participatory approach that allowed communities to identify their own water and sanitation problems and determine their own priorities for intervention. The intervention recognized that the "traditional" approach of providing safe drinking water and sanitation facilities to every household and school was not feasible due to various factors such as limited resources, logistical challenges, and the remote nature of some communities. They termed an approach as being traditional if it constituted the direct provision of hardware and software to participants.

Accordingly, the intervention employed a community-led approach that involved identifying the most critical areas where water and sanitation interventions were needed and prioritizing them based on their level of need. This approach involved training local masons to construct water and sanitation facilities using appropriate technologies and materials and engaging community members in digging latrine pits, carrying materials, and providing labor during the construction process. The VEA intervention also used appropriate technologies and materials that were available and affordable within the communities, and hence actively sought to optimize the engagement of community members in the planning, implementation, and maintenance of water and sanitation facilities. The intervention also employed a behavior change approach that focused on promoting good hygiene practices and implemented a verification process to ensure that the communities were using the facilities properly and maintaining them over time. Once a community achieved a defecation-free status, they received a certificate that recognized their achievement, promoting healthy competition between communities.
Using such an approach, the VEA intervention attempted to ensure that the interventions were tailored to the local context and needs of the communities, and that the communities were invested in and actively took ownership of the interventions. Furthermore, by involving community members in the planning and implementation process, the VEA intervention was able to better identify and overcome logistical and cultural barriers that might have hindered the success of the intervention.

As for the hardware made available, and apart from the latrines and their construction equipment, the sanitation equipment also encompassed tippy-taps, which were to be mounted near latrines and food handling areas, as well as a sanitary scoops designed specifically for removing feces. Plastic potties were also provided to aid in potty training the infants and transfer of infant feces to latrines.
Since the study simply mentioned providing "tippy-taps" without describing them, external literature was consulted to do so: tippy-taps are a low-cost and simple handwashing device that is widely used in low-income and rural areas where access to running water and soap is limited. The device consists of a plastic or metal container, such as a jerry can or a plastic bottle, suspended from a stick or a branch, with a rope or string tied to the container's cap. The rope or string is then tied to a foot pedal or stick, which is used to tip the container to release water when pressed with a foot (UNICEF, 2019). Tippy taps are designed to be a hands-free device, so the person can wash their hands without touching the container or the tap, reducing the risk of contamination (WHO, 2009). The devices are widely used in developing countries, particularly in rural areas and schools, to promote good hygiene practices and prevent the spread of diseases such as diarrhea, cholera, and COVID-19 (TeachAids, 2020).

4.2 OUTCOMES MEASUREMENT

The evaluation procedure and sequence employed to assess the outcomes of an intervention along its causal chain were similar across studies. The baseline conditions were first recorded, the intervention was then set up and initiated, and finally the functionality as well as the primary and secondary outcomes of the intervention were measured to be finally analyzed. The following section aims to elaborate on the targeted outcomes that the trials investigated, along with their employed data acquisition method.

4.2.1 BASELINE CONDITIONS

Baseline conditions data were mainly acquired through questionnaires presented to the recipients at the beginning of the trials. In general, the questionnaires collected information related to demographics, socio-economic characteristics, water, hygiene and sanitation practices, such as time spent collecting water and transported volume, drinking water sources, water treatment practices, infant feeding practices, water, sanitation, infant water exposures, any recent illness in the enrolled participants (e.g., diarrhea, fever, and cough), or any recent health facility visits for illness. Furthermore, in several studies (Crump et al., 2005; Boisson et al., 2009; Boisson et al., 2010; Albert et al, 2010; Stauber et al., 2012; Gunther et al., 2013), the initial quality of the Point of Source (POS) and Point of Use (POU) water of the recipients was tested at the beginning of the study. Only one study, (Morris et al., 2018), investigating the effectiveness of silver colloidal ceramic water filters on diarrheal reduction and lowering of *Cryptosporidium* contraction, tested for both the water quality of the POU household water via *E. Coli* concentrations, along with the free residual chlorine levels to test for the current water treatment practices of the households who reported to treat their water with chlorination.

However, in certain studies in the dataset, some relevant baseline conditions that would allow the establishment of a causal relationship between the effect of the intervention and improvement in the targeted outcome were not collected. For example, in (Garrett et al. 2008), a trial that was looking into the effectiveness of different interventions, including safe water storage with chlorination, promotion of rainwater harvesting and the implementation of latrines, on diarrheal reduction, did not collect diarrheal episodes data at baseline. This is especially problematic in case there were a difference in baseline occurrences between control and intervention groups, thereby inducing bias.

Finally, there was some absence of data separation based on baseline gender and social identifiers, as 65% of the studies disaggregated their data by gender and only 34% by social identifiers, such as the level of education attained, household income or employment status.

4.2.2 HEALTH OUTCOMES

The most frequently investigated outcomes by the trials in the dataset were health associated. In all of the trials that explored health-related indicators, the main targeted population was children, since they were deemed to be the most vulnerable population category experiencing a high rate of morbidity and mortality due to contracted

diseases from contaminated water. The health outcomes that were explored were: diarrhea, stunting and anemia and Acute Respiratory Infections (ARI's).

Actually, upon further research in literature outside of the dataset, anther motivation for applying this approach may be hypothesized as all of these conditions were found to be interconnected. Diarrhea, stunting, and anemia are important indicators of malnutrition and poor health in children. Diarrhea is a condition in which a child experiences frequent watery bowel movements, leading to dehydration and nutrient loss (Haider et al., 2019). Stunting on the other hand, refers to impaired growth and development caused by chronic malnutrition, while anemia is a condition characterized by a lack of iron in the blood, leading to fatigue, weakness, and impaired cognitive function (Kosek et al., 2017). These indicators are closely linked, as they can all result from inadequate nutrient intake and poor hygiene practices. Additionally, ARI's can further exacerbate malnutrition and subsequent stunting, and weaken the immune system, leading to more frequent infections and a higher risk of mortality (Nair et al., 2020). It is hence why certain interventions, such as in the SHINE trial or the WASH Benefits Kenya trial, addressed malnutrition along with improving water quality and hygiene practices.

A. DIARRHEA

FREQUENT APPROACH

Diarrhea was by far the most recurrent health outcome measured. It was always stated by the trials as being one of the leading causes of child mortality in SSA.

In 19 trials in the dataset, diarrhea was the primary outcome investigated. All trials, except two (Garrett et al., 2008; Patel et al., 2012), clearly defined what a diarrheal episode is, stated according to the WHO definition, which goes as follows: the passage of three or more loose or liquid stools per day (or more frequent passage than is normal for the individual). The WHO also defined a diarrheal event to be deemed as a new episode if the child lasted at least three days without any diarrheal symptoms.

To measure its occurrence, several methods were employed. The most common method utilized was collecting self-reported incidences of diarrhea from the caregivers of children aged a maximum of 5 years. In 16 of the 19 trials, data was collected by field workers on household visits, whereby they collected the reported number of incidents that happened in the past 7 or 14 days from the caregivers. The frequency of the visits ranged between one and two weeks, and so the recall time of incidences was 7 or 14 days.

TRIALS WITH DISTINCTIVE APPROACHES

A few trials in the dataset made use of design setups that differed from abovementioned approach of collecting participants-reported diarrhea data through timely household visits. In the following, the differing approached are detailed.

1. Use of Pictorial Diaries: Two trials, (du Preez, 2008; du Preez, 2011), investigating the effectiveness of ceramic filters and solar disinfection in reducing diarrheal incidence respectively, opted to shy away from reporting diarrhea incidences through recall, and attempted to collect more accurate temporal data. In these trials, caregivers were given pictorial diaries to denote the daily diarrheal (non)incidence. These trials were particularly distinguished since they were the only trials in the dataset that were able to differentiate between two types of diarrhea: dysentery and

non-dysentery, by instructing participants to distinguish between loose watery stools and stools including blood or mucus for don-dysentery and dysentery diarrhea respectively. Both trials made use of the daily reporting of diarrhea and hence specified a diarrheal event simply as a day on which diarrheal symptoms were observed.

As for the distinction between the two types of diarrhea, (du Preez et al., 2008) reports that dysentery diarrhea is a severe form of the illness. It is typically caused by bacteria such as *Shigella*, *Salmonella*, or *E. Coli*. Children with dysentery diarrhea usually have frequent, bloody bowel movements, accompanied by abdominal pain, cramping, and fever. In severe cases, dysentery can lead to complications such as dehydration, malnutrition, and even death.

Non-dysentery diarrhea on the other hand, is usually caused by parasitic infections such as *Giardia*, *E. Coli* or *Cryptosporidium*. These infections are also spread through contaminated food and water and can lead to mild to moderate diarrhea in children. Children with non-dysentery diarrhea may have loose, watery stools, mild abdominal discomfort, and possibly fever, but they are less likely to have bloody stools than those with dysentery.

Through these trials, the researchers attempted to gain more insight on the manifestation of *Shigella* infections, especially that they report that information on this type of infection in Africa is rather limited. Furthermore, since the two types of diarrhea are caused by different infectious pathogens, they may need different interventions to prevent and treat them. The trials hence attempted to evaluate the effectiveness of popular POU interventions in targeting either types.

2. Testing Stool Samples for Infections: Instead of indirectly inferring the contribution of pathogens to diarrheal occurrences via direct observations of diarrhea episodes, two trials directly tested the presence of pathogens that usually cause diarrhea in stools of children under two years old: *Cryptosporidium* in (Morris et al., 2018) and Environmental Enteric Dysfunction (EED) in (Gough et al., 2020) of the SHINE trial The testing of those pathogens was accompanied with household water quality testing.

3. Collecting the Number of Visits to Healthcare Facilities: Opting for a more measurable indicator of diarrheal occurrences outcome, one study, (Morris et al., 2018), used changes in the number of healthcare visits with data extracted from logs at the centers. By checking for the changes in the frequency of visits, the study aimed to note not only the occurrence of diarrhea episodes, but also their severity, and the potential economic gain from reduced healthcare expenses.

B. STUNTING

The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) trial is motivated by the premise that environmental enteric dysfunction (EED) is a major underlying cause of both stunting and anemia, that chronic inflammation is the central characteristic of EED mediating these adverse effects, and that EED is primarily caused by high fecal ingestion due to living in conditions of poor water, sanitation, and hygiene (WASH). Environmental enteric dysfunction (EED) may be an important modifiable cause of child stunting.

C. ACUTE RESPIRATORY INFECTIONS (ARI's)

Only two trials in the dataset, (Patel et al., 2012; Swarthout et al., 2020) both set in Kenya, explored the effectiveness of combined WASH and nutrition, as well as hygiene messaging in combination with the installation of water points, on the reduction of ARI's respectively. As stated above, nutritional supplements have been an intervention utilized

to strengthen the immune system of children and reduce their susceptibility to contract infections. Furthermore, ARI's could be developed due to respiratory infections gestated from fecal-oral transmission (Swarthout et al., 2020). This is the reason why interventions to reduce ARI's opt for attempting to block the transmission pathways of fecal contaminants via sanitation and hygiene hardware and software interventions an or provide access to safe water.

The prevalence of ARI's was defined differently in the two trials and had different targeted age groups. According to (Patel et al., 2012), an ARI is considered to occur "when fever and cough or difficulty breathing in the preceding 24 hours" for children less than 5 years. On the other hand, (Swarthout et al., 2020) defined an ARI to be symptomatic of having "cough or difficulty breathing, including panting or wheezing, within 7 days before the interview—in children younger than 3 years". Occurrence data was collected biweekly (Patel et al., 2012) and weekly (Swarthout et al., 2020) by field workers on house visits.

4.2.3 BEHAVIORAL OUTCOMES

The degree of the acceptability and uptake of an intervention by the recipients were measured via different direct and indirect indicators.

A. DIRECT OBSERVATIONS

Direct indicators involved collecting the amount of consumed product during house visits, such as the number of empty chlorination bottles in (Garrett et al., 2008; Albert et al., 2010; Mengistie et al., 2013; Solomon et al., 2020), or the number of opened flocculent disinfectant sachets (Crump et al., 2005) at unannounced visits. *E. Coli* and turbidity were also sometimes tested in effluent and stored water to test for intervention utilization (Luoto et al, 2014) at unannounced visits as well. Usage was also measured through self-reported practices by the recipients (Albert et al., 2010; Stauber et al., 2012; Cha et al., 2015; Luoto et al., 2014; Boisson et al., 2010; Morris et al., 2018; Quattrochi et al., 2021). One study, (Boisson et al., 2009), estimated utilization by analyzing the responses of participants regarding their experience or satisfaction levels with the product.

DEFINITION OF UPTAKE

The definition of what interventions considered an indicator of hardware utilization and what they termed as "uptake" varied widely across the studies and the differentiations between the two were not always made clear.

Some definitions of hardware utilization used as indication the utilization of the intervention at least once in the past week or on the day prior to the household visits. Others defined it as the total number of days the hardware was used between visits, those visits having a frequency that ranged between one week or two.

Likewise, when presenting the results of "uptake" rates, some studies were not clear on the parameter they were reporting. This could best be demonstrated by (Rosa et al., 2014) studying the effectiveness of hollow fiber filters in improving the water quality of the household in Rwanda. At the end of the study, the researchers reported an 89.2% uptake rates, while later on stating that 25% of households were consuming water from an untreated source within one of the five household visits. The two different definitions of uptake were implicitly differentiated: one having received and used the filter at least in one of the five household visits, and the second is having consistently used the filter throughout the duration of the intervention.

While in (Rosa et al., 2014) the distinction was established, in other studies, it was not made clear. In such studies, the indication of uptake on the day of the household visits would have been defined at the beginning of the study but their definition of uptake including the rate or frequency of utilization were not.

B. INDIRECT OBSERVATIONS

Indirect observations on the other hand, were varied in nature. Some trials inferred the extent of usage through observing physical signs of product functioning, such as the presence of soap in sanitation facilities, spot checks improved sanitation (Kremer e al., 2012; Fagerli et al., 2020; Solomon et al., 2020)). Improved sanitation indicated that the primary place of defecation was a ventilation-improved pit latrine, a pit latrine with slab or a composting latrine, or if flushes/pours were directed to a piped sewer, septic tank or pit latrine.

For POU water treatment interventions, several studies tested free chlorine residual on-site to determine whether households had treated their water, included stored water (Garrett et al., 2008; Bossion et al., 2009; Mengistie et al., 2013; Luoto et al., 2014; Cha et al., 2015; Solomon et al., 2020; Fagerli et al., 2020), or if they had performed additional treatment on filtered water (Morris et al., 2018). In case of filtration, if a filter was found hung for use with water in the top vessel of the device, then the filter was deemed to have been used (Boisson et al., 2010).

Using self-reported data was often stated as a study limitation since they may generate biased findings and threaten the validity of the studies (Boisson et al., 2009; du Preez et al., 2011; Mangestie et al., 2013; Stewart, 2019). Reporting courtesy is based on social desirability bias whereby participants may be more likely to report what they think the researchers want to hear or what is socially acceptable rather than what actually occurred. Moreover, certain studies warned about taking acceptance of uptake rates as accurate since participants may have undergone the Hawthorne effect which leads to positive adherence because they were conscious that their behavior was monitored (Fagerli et al., 2020; Solomon et al., 2020).

All in all, These evaluation practices, while being practical and optimal especially with budget considerations, could be particularly flaky since the continual use of the intervention was not monitored.

DISTINCTIVE APPROACHES

That being said, (Rosa et al., 2014) was the only study within the dataset that had a different approach on usage monitoring. The study is a pilot trial for a program that aims to enhance household water quality and air quality by handing out a hollow-fiber LifeStraw and an improved cook stove respectively. (Rosa et al., 2014) was a 5-months trial and featured a sub-trial that ran along it, detailed in (Thomas et al., 2013), whereby 23 sensor-mounted filters where handed out every two weeks to randomly selected households. The surveyed households amounted to approximately 40% at the end of the trial with daily filter use records for 2 weeks each. The production and installation of each sensor amounted to around USD $500. By comparing the data from the sensors to that of the surveys administered to the participants at each household visit (every 2 weeks), it was shown that the use of the filters was 36% less than what is reported. It was also shown that between 50 and 80% of households were not treating their minimum needed daily volume of water, further suggesting that they may be utilizing another, possibly contaminated, source. However, it is worth noting that some sensors had technical defects that were spotted during

the study and that may have influenced the findings. The number of sensors that experienced these challenges and their repercussions were not reported in the study.

4.2.4 TECHNICAL OUTCOMES

How well an intervention was operating as intended was measured throughout the trials.

1. Water Supply Interventions: For an intervention focusing on water supply, in both cases in the dataset (Quattrochi et al., 2021; Cha et al., 2015), a water source was deemed as improved if it proved to be a functional water point with sufficient water flow. Therefore, the main outcome measured to indicate functionality was water point flow rate.

2. Water Treatment Interventions: For an intervention focusing on water treatment, its functionality, indicating that the treatment was operational and microbiologically effective, was measured via testing for the parameters listed hereafter along with their indicative safety benchmark values. These values were always based on the WHO recommendations of values for water safety:

1. Fecal contamination: more than 1 Colony-Forming Unit (CFU) of *Escherichia Coli* (*E. coli)* per 100 mL of water;

2. Total Coliforms: no more than 0 CFU of total coliforms per 100 mL of water. However, one study, (Gunther et al., 2013) used a higher threshold for contamination of 1000 CFU/mL;

3. Turbidity: turbidity was only measured in case it was tested for during baseline data collection and was found high; its indication was that water should have a turbidity of less than 5 Nephelometric Turbidity Units (NTU).

In the only trial in the dataset that assessed the impact of household ceramic filters on diarrhea (Morris et al., 2018), cryptosporidiosis prevention, and water quality in rural western Kenya, the water was tested for both *Cryptosporidium* concentration and fecal contamination.

3. Interventions Including Health Messaging: For the accompanying software interventions such as health and hygiene messaging, recommendations on best-practices and intervention hardware handling, only one study, (Fagerli et al., 2020) evaluated the degree of communicability and clearness of the practice guidance they offered when handing out ceramic filters to intervention households. They did so via questionnaires featuring multiple choice questions at the end of the trial.

4.3 TRIAL DESIGN AND SETUP

4.3.1 TRIAL LOCATION AND DURATION

A. TRIAL LOCATION AND SETTINGS

TRIAL LOCATIONS

The studies in the dataset were distributed across nine countries, with Kenya having the highest number of publications. Three trials in the dataset had several publications relating to sub-trials of the same study. These studies are: the SHINE trial having six publications relating to it, WASH Benefits Kenya trial having three and (Albert et al., 2010; Luoto et al., 2014) revolving around two complementary sub-trials of CARE Kenya trial. The details of the main trials are detailed in Appendix A.

The graph in Figure 14 depicts the geographical distribution of the trials in the dataset. One thing to note is that the publications belonging to same study were grouped into one.

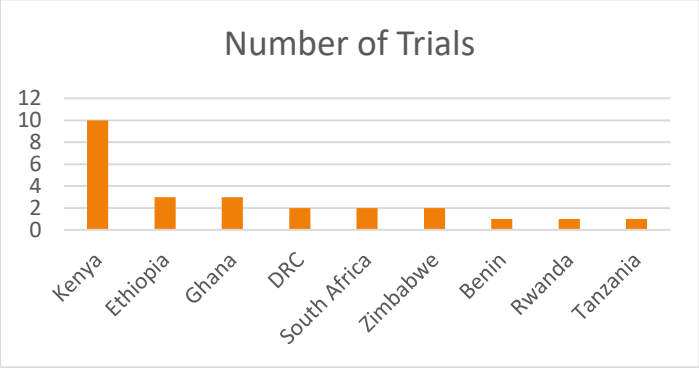

Figure 14: Number of Trials/ SSA Country in the Dataset

TRIAL SETTINGS

Furthermore, the settings of several trials in the dataset were especially targeted. Essentially, the pre-existing baseline settings were in a lot of studies appointed conditions for inclusion. This is especially true for environmental, infrastructural or health parameters. For example, (Boisson et al., 2010) chose areas with water sources high with fecal contamination, (Fagerli, 2020) chose a province that has the highest rate of infant mortality and exceptionally high diarrheal prevalence. (du Preez, 2008; Quattrochi, 2021; Cha et al., 2015) also targeted areas with an excessive lack of amenities such as sanitation or water coverage.

B. TRIAL DURATIONS

The duration of the trials varied greatly. Two trials, (Cha et al., 2015; Garrett et al., 2018), spanned a duration of two months each and were the shortest studies in the dataset. On the other hand, only four trials (Boisson et al., 2010; du Preez et al., 2011; SHINE; Swarthout et al., 2020) were longer than a year with the SHINE trial the longest trial spanning 29 months and being the only one to last more than a year. All in all, 48% of the trials

were shorter than 6 months, 39% between 6 and 12 months, and 13% more than 12 months. The duration of a trial has been stated as a factor influencing the findings and their generalizability of several studies in the dataset.

1. User Uptake: Firstly, short duration trials have been reported to have the potential to provide a limited representation of the impact of an intervention. (du Preez et al., 2008), when reporting the results of their 17 months-trial assessing the effectiveness of solar disinfection and health messaging in decreasing diarrheal occurrence among children younger than 5 years, emphasized the decrease in user adherence over the course of the trial, and hypothesized this trend to be due to the decrease in household visits for health messaging and reminders in the second half of the trial.

2. Microbiological and Technical Performance: Secondly, (Fagerli et al., 2020) observed a decline in the microbiological performance of the LifeStraw hollow fiber filter used in their five months-trial. Also, over the trial period, 25% of the filters were broken and needed either repairing or replacement. The study emphasized the need to conduct a longer period trial to fully observe the performance of the filters over time.

Another study, (Morris et al, 2018) conducted a short trial of 4.5 months in rural Kenya whereby ceramic filters were evaluated for the decrease of diarrheal incidence and *Cryptosporidiosis* prevalence. During the trial, 10% of the filters were broken or temporarily out of service. In the study, authors concurrently present another study that is not part of this study's dataset, (Clasen et al., 2015), that investigated the performance of ceramic filters in similar settings. The impact evaluation found that over 30% of the filters broke down and needed fixing or replacement. The authors hence emphasize the need for a supplementary investigation of the long-term performance and durability (uptake) of the filters.

3. Seasonal Variations: Seasonal variations and their consequent effects on parameters that influence the performance and uptake of an intervention were also limitations mentioned in several studies in the dataset. For instance, (Albert et al., 2010) was a six-months study in rural Kenya that investigated end-user preference between three POU household water treatments. The authors noted that the choice and usage of the treatments varied as the season changed. During the rainy season, households that did not use harvested rainwater and relied on surface water employed more the received treatment since the surface water was perceived as dirty due to its turbidity.

Another trial that lasted for only two months and which was one of the two shortest study trials in the dataset, (Garrett et al., 2008) studied the effect of chlorination and safe water storage on diarrheal prevalence in rural Kenya. The study noted that since the trial was conducted in the rainy season when waterborne pathogens are at their highest, its findings, that showed chlorination to be the only effective intervention against diarrhea with a 43% reduction, may be biased.

All in all, the trials emphasize the inability of short timed trials to account for and simulate the long term sustainability of an intervention both in terms of technical performance as well as recipient usage. One phenomenon that was not mentioned in the dataset but is nonetheless worthy to note is that WASH interventions often involve promoting behavioral change, which may require a longer time period than the duration of a short-term trial to become established and sustainable (Wolf et al., 2014). This further reinforces the concerns over the inability of short trials to adequately account for the different phenomena that occur throughout and after the implementation of the intervention.

4.3.2 TRIAL DESIGN

A. RANDOMIZATION

The dataset comprised of only publications related to clinical trials. All of them had a randomized controlled trial design, which may signify the degree to which the research and academic funding sector favors this method of research.

DEFINITION OF RANDOMIZATION

Consulting external literature, (Rothwell, 2005), indicated that in RCT's, participants are randomly assigned to either an intervention or a control group. The control group receives a placebo or no treatment, allowing researchers to compare the outcomes of the intervention group with those of the control group. By randomly assigning participants to groups, RCT's aim to minimize the influence of factors that could bias the results, such as differences in baseline conditions including among others, age, gender, health status and indicators, point of source (POS), level of education and socio-economic status. Thus, the term "controlled" refers to the ways in which RCTs aim to control the influence of extraneous factors on the outcomes of the study.

RANDOMIZATION IN THE DATASET

Although the trials were intended to be randomized, several of them were unsuccessful in doing so properly. First off, (Stauber et al., 2012) who studied the effectiveness of biosand filters in reducing diarrhea in rural Ghana, initially intended to do a randomization at household-level but were obliged to do it at village-level since the elders of the selected villages did not accept to have a household-level randomization. The reason behind their refusal was not mentioned. The study reported the randomization scale-up as a limitation only because it reduces the sample size.

Secondly, some studies have especially targeted or handpicked their participants. One study, (Cha et al., 2015), studied the effectiveness of improved water supply through borehole construction or rehabilitation and the provision of latrines, on the prevalence of diarrhea at children aged 5 years or younger. This study only included in its per-protocol analysis the households that conformed to the ascribed protocol and guidelines while those who did not were excluded. Another study, (Solomon et al., 2020), studied diarrheal occurrence change among children under the age of five years due to chlorination. The study excluded households with severely ill caregivers due to their potential inability to respond to questionnaires. This study also excluded households with children less than six months of age with incessant diarrhea, but offered no justification for this exclusion criteria. The study also did not report the number of households that it dismissed due to the aforementioned criteria, or mentioned these criteria as a study limitation.

Thirdly, the pre-existing baseline settings were in a lot of studies appointed conditions for inclusion. This is especially true for environmental, infrastructural or health parameters. For example, (Boisson et al., 2010) chose areas with water sources high with fecal contamination, and (Fagerli, 2020) chose a province that has the highest rate of infant mortality and exceptionally high diarrheal prevalence. Furthermore, (du Perez, 2008; VEA; Cha et al., 2015; Rosa et al., 2014) all chose areas with an excessive lack of amenities such as sanitation or water coverage. All of the studies, however, acknowledged these selection as potential limitations, stating that their selection choice

could compromise the generalizability of their findings to other populations. (du Preez et al., 2008) further mentioned that their selection could be the cause of the high reduction in diarrhea observed. However, none went into further details about the additional potential ensuing effects.

B. BLINDING

DEFINTION OF BLINDING

Another trial characteristic is blinding, which is a technique used in clinical trials to reduce bias by preventing participants and/or researchers from knowing which group participants are assigned to. Referring to external literature for the official definition of the act of blinding in trials, blinded studies are classified as single-blind or double-blind depending on the degree of blinding. In single-blind studies, participants are unaware of their group assignment, while in double-blind studies, both participants and researchers are unaware. Blinding can help to reduce bias, particularly the placebo effect, where the mere act of receiving treatment can have a positive effect on the outcome (Moher et al., 2010). Non-blinded studies, on the other hand, do not use blinding techniques and participants and/or researchers are aware of the group assignments (Devereaux et al., 2001).

BLINDING IN THE DATASET

In the dataset, all trials except for (Boisson et al., 2010; WASH Benefits Kenya), were non-blinded RCT's. The WASH Benefits Kenya trial was an RCT with a single-blind whereby only the researchers were blinded. The second study was (Boisson et al., 2010), that consisted of a randomized double-blinded placebo-controlled trial. It assessed the field performance, use and effectiveness of the hollow fiber filter in preventing diarrhea in 240 households in rural DRC. This study was particularly special as it was the only study in the dataset that implemented a placebo through distributing ineffective filters to the control group. The placebo was made possible because the POS that 86% of the participants utilized was water with low turbidity. Compared to the active filters, the placebo filters looked and weighed identically o their counterpart, had the same outer components and produced the same flow rate. The only difference is that the membranes were replaced by additional piping to mimic the weight and flow rate of the active filters. the blinding, however, was not effective as a considerable number of the control group members identified their filter as active. This is due to the fact that the placebo turned out to be microbiologically effective and was able to eliminate up to 90% of fecal coliforms. The researchers do not know exactly the reason behind this phenomenon but hypothesized that a biofilm layer formed on the inside of the plastic pipe installed to mimic the membrane in the active filter, and was able to retain bacteria and particulate matter.

All in all, blinding was always listed as one of the limitations of the studies. The main reason that was given repeatedly by the publications as to why they did not implement one is that treatment inactivity is hard to conceal. This was reported to be especially true for POU chlorination interventions or if the POS water was turbid, or if the intervention evaluation revolved around the construction or rehabilitation of boreholes, since inactive boreholes are easily detected. In fact, the abovementioned trial conducted by (Boisson et al., 2010) was a highly cited publication, both within the publications in the dataset and externally, often mentioned to justify the lack of placebo implementation and emphasizing that designing a neutral placebo proved to be a challenging task.

Only one study, (Stauber et al., 2012), mentioned the lack of introducing a placebo filter was for ethical considerations, essentially since caregivers were reported to filter their water by boiling it when for their child's

consumption. The authors hence hypothesized that the ineffective filter would be used for child consumption and expose the recipient to potential harm.

4.3.4 FIELD CHALLENGES

Only three studies reported logistical challenges that occurred while conducting the trials.

1. (Morris et al., 2018): The study, when investigating the ability of ceramic filters to decrease diarrheal incidence and *Cryptosporidium* prevalence, faced logistical challenges in consistently collecting infant stool samples: when diarrhea was reported and the infant did not produce a stool during the visit, caretakers were asked to collect a sample of the next stool for testing. However, returning to the home to collect the sample before the next weekly visit proved to be challenging. The authors nonetheless did not report on the number of missed samples or on the repercussions that this might have imposed.

2. **(Quattrochi et al., 2021), VEA Trial:** The second study, VEA, was a large scale program in the DRC investigating the effects of a community-driven water, sanitation and hygiene intervention on water and sanitation infrastructure, access, behavior, and governance (detailed in section 4.1), also faced challenges in collecting health outcomes indicators, especially stool samples. This was mainly due to budgetary constraints limiting the number of recruited personnel that were needed to complete the task, especially with the widespread reach of the intervention. This drawback was listed as a study limitation but was not elaborated on in terms of the amount of missing data and how they may have affected the observed outcomes.

3. (Christensen et al., 2015), WASH Kenya Trial: Finally, in a pilot study done for the WASH Benefits Kenya trial by (Christensen et al., 2015) whereby the amelioration of existing latrines proved to be a difficult and time-consuming endeavor. Field workers had to be taught about how to uniformly assess the status of the hardware at baseline to have other enumerators come in with the needed equipment and parts after 2 months. However, what was noticed is that the status of the latrines would have changed and needed different equipment or that the new enumerators had a different assessment of the condition of the hardware. Also, commissioning locals to do the reconstruction required a substantial amount of time and disbursement. As a result, the main intervention only provided the construction of latrines in case one was not present.

Moreover, in this trial, the promoters ended up making approximately six visits per month on average during the entire trial period, while in the Kenyan study, promoters made an average of one visit every 1-2 months. The reasons behind these disparities despite the initial trial design were never reported.

4.4 STUDY RESULTS

The results of the trials are presented along with some general trends observed.

4.4.1 HEALTH OUTCOMES

A. DIARRHEA

When it comes to the reduction of diarrheal prevalence, the results within the dataset were mixed. Twelve trials obtained positive results; they revolved around chlorination (Garrett et al., 2008; Mengistie et al., 2013; Solomon et al., 2020), filtration (du Preez et al., 2008; Boisson et al., 2009; Stauber et al., 2012), flocculent disinfection (Crump et al., 2005), solar disinfection (du Preez et al., 2011), storage (Gunther et al., 2013), either of the three types POU treatments (chlorination, filtration and flocculent disinfectant) combined with messaging (Garrett et al., 2008), and finally rehabilitating boreholes and latrines combined with messaging (Luoto et al., 2014).

In contrast, six other trials observed no effect on diarrheal prevalence while using the following interventions: filtration (Boisson et al., 2010; Morris et al., 2018; Fagerli et al., 2020), combined WASH and nutritional supplements (SHINE; WASH Benefits Kenya) and lastly hygiene and health messaging combined with the provision of sanitation hardware in (VEA).

However, one thing to not is that the indication of success in diarrhea reduction and the associated baseline reference values, or what is deemed as effective or enough, were not established. An intervention was deemed as successful in different trials, with success rates ranging between 11% to 60%.

Several observations can be made from the distribution of the effectiveness results. First off, all three large scale trials in the dataset have observed null effects on diarrhea reduction. Furthermore, all of the trials with null effects are long-term trials, lasting one year or more. Lastly, the only double-blinded trial in the dataset, (Boisson et al., 2010), a trial design hailed for its minimization of bias, the diarrhea reduction was also unobserved.

Furthermore, no observable pattern was detected linking the results on diarrheal reduction and the type of intervention used except for chlorination not being featured in any of the studies that had null effects on diarrhea.

B. OTHER HEALTH OUTCOMES

ARI's

As previously stated, only two trials, both performed in Kenya, investigated changes in the prevalence of ARI's, one was a sub-study of the WASH Benefits Kenya trial (Swarthout et al., 2020) which lasted for 20 months and the other, (Patel el al., 2012) was a 12-months trial whereby a hygiene and handwashing curriculum and new water points were installed in schools. Both trials start with a hypothesis reported to be backed by considerable evidence in literature, which states that improved handwashing leads to a decrease in ARI's. However, the studies resulted in contradicting results as (Swarthout et al., 2020) obtained null effects on ARI's incidences for all the intervention arms while (Patel et al., 2012) observed a reduction in the median percentage of participants with ARI's.

Apart from having different types of interventions, one of the main differences between the two trials is the degree of uptake of the interventions by the participants. In (Patel el al., 2012), an improvement in handwashing practices was observed for both the children and their caregiver. In (Swarthout et al., 2020) on the other hand, even the WASH intervention arms that included sanitation improvement, mainly the limiting of fecal contamination via the construction of latrines, and those providing soap and tippy taps for improved handwashing practices, were not effective in limiting ARI's.

These findings may reinforce the hypothesis that rigorous behavioral promotion coupled with high intervention uptake are required to effectively instigate clinically-relevant decreases in ARI's.

STUNTING, EED AND ENTERIC INFECTIONS

In the SHINE trial, no impact on EED or in Enteric Infections was observed in any of the intervention arms, and no effect that translated into reduced stunting. However, the trial did find that improved nutrition, with and without basic WASH interventions, did improve anemia as well as stunting, which was measured through anthropometric measures.

In WASH Benefits trial that was performed in Kenya and Bangladesh, small improvements in child growth and motor skills leading to decreased stunting. Although the WASH Benefits Kenya trial in general was hailed by other literature, both within the dataset and beyond it, for having an intensive and frequent contact between field operators and the households, the Bangladesh trial was able to provide more home visits than the 2 times a month for 2 years in the original trial design. The rate of uptake was also slightly higher in Bangladesh than in Kenya. The researchers therefore hypothesized that the frequent visits enabled the development of a bond between a peer from the community and the caregivers, providing support against several stressors especially during the first period of a child's development, and sustaining the new mothers with mental health support, and the space to be able to actually uptake new practices.

In general, both trials concluded by emphasizing the need for more advanced interventions than the basic offerings of the trials.

4.4.2 BEHAVIORAL OUTCOMES

DIFFICULTY IN REPOTING THE RESULTS

In this study, reporting on the results of the degree of uptake of the interventions by the recipients in the trials of the dataset proved to be a challenging task. The main reason behind this is that the definition of uptake varied widely across studies and the reporting of the findings was sometimes vague, which yielded to difficulties in establishing comparability between study results and their interpretation. That is why amassing the general uptake rates of the trials was not attempted. Instead, this study reports on the preferences that trial participants expressed in the few trials in the dataset that presented its subjects with multiple intervention options. This study also reports on the reasons behind these preferences and any additional behaviorally relevant observations made by any publication within the dataset.

USER PREFRENCE AND BEHAVIORAL PATTERNS

1. (Boisson et al., 2009): Firstly, (Boisson et al., 2009), was one of the few studies to record feedback on user experience and encountered problems. The trial investigated the effectiveness of the portable version of the LifeStraw hollow-fiber filter. This particular study reported the lowest usage rate in the dataset amounting to only 13% of the recipients, and that is despite the filter being handed out for free and the recurrent household visits to record diarrheal prevalence during which renewal messaging about the importance of water safety is done. This low level of uptake was found to be due to the impracticality of using the filter, with 73% of respondents reporting the force needed to suck on the filter to be able to draw water. In addition, 42% found the filter to be too slow in generating consumable water while 30% found it to be too big to be practical for it to be carried around.

2. CARE Kenya sub-Trials: Secondly, (Albert et al., 2010) was one of the few trials that closely looked into end-user preference. The trial was conducted in two areas: rural Kenya and urban slum Bangladesh, but this study focused only on the results of the Kenya trial. The study presented its participants with several POU household water treatment options: chlorination bottles, flocculent-disinfectants and ceramic filers. All three treatments were rotated onto the households and the preference of the participants and the reasons behind them were explored. The study found that participants favored treatments in the following order of preference: filters, flocculant disinfectants and chlorination.

The study also found that ease of use and the speed in generating usable water were the two most valued treatment characteristics. These were followed by valuing water without unpleasant odor and taste, such as those resulting from chlorination. The study especially noted that despite illustrating to households the functions and importance of the treatment, they were still suspicious towards it due to induced disagreeable water. They were also suspicious of it since babies under the age of 3 years were reported to having had a hard time digesting the water and crying whenever it was consumed due to what caregivers presumed as stomach aches. The affordability of an intervention was surprisingly ranked as a fourth preference parameter, which suggests that end-users may be more likely to buy household water treatment options if they are efficient and user-friendly, rather than if they are just affordable. Finally, the last two characteristics by order of preference were respectively durability of hardware and microbiological performance.

These findings may reveal the possibility that the trial was not successful enough in communicating the safety and importance of the treatments, revealed by the suspicion around chlorination. It might also have not been entirely successful in explaining the risks of unsafe water and the health benefits and importance of clean water. Furthermore, such preferences and priorities of intervention utilization were repeatedly mentioned across several studies as reasons behind the degree of uptake of the interventions.

The complementary study of the aforementioned one, (Luoto et al., 2014), investigating the type of behavioral messages that increase the uptake of the interventions, discussed the findings of the Kenyan and Bangladesh trials. The findings from separate RCT's conducted in two different settings were meant to increase confidence in the results. The study found that providing free safe water products along with repeated educational messages increased the use of safe water treatment in both countries compared to a control group or the baseline. However, there were significant differences in the rate of product usage between Kenya and Bangladesh, as participant in Kenya self-reported usage three to four times more than participants in Bangladesh.

In Kenya, framing safe water technologies as both preventing diseases and improving health resulted in an increased in usage rates, as compared to the use of a positive frame that only mentioned improving health. Similarly, marketing messages that asked consumers to publicly commit to water treatment and provided a reminder poster had almost the same impact. Results from Bangladesh were similar, with the commitment with reminder treatment resulting in a higher increase in water treatment as compared to the framing treatment.

The experiments however were not designed to explain the differential take-up or attitudes across settings. Yet the authors speculated about several possible reasons. For instance, in Kenya, there had been more social marketing of the products, which could have created a stronger local perception that it is socially acceptable to use them. Self-reported usage rates were substantially higher in Kenya, but slightly lower in Dhaka, indicating possibly significant differences in courtesy bias across the two settings. Moreover, more households in Kenya relied on turbid surface water during the study period, and more Kenyan households reported recent diarrhea at the baseline. Additionally, households in Kenya were provided with buckets with taps along with their free water treatment products, which were worth approximately a day's wage for an informal laborer, and this could have contributed to the higher take-up rates in Kenya.

In Dhaka, a real-money auction was conducted after the final survey round at both treatment and control homes, where treatment households showed a lower willingness to pay for chlorine-based products compared to control households, and there was no increase in willingness to pay for the filter. This indicates that most households in Dhaka did not prefer the chlorine products.

3. (Swarthout et al., 2020): Another trial that noted a behavioral pattern was (Swarthout et al., 2020), a sub-trial of the WASH Benefits Kenya trial investigating the effects of nutritional supplementation with or without simple WASH interventions on the prevalence of ARI's in Kenya. Their evaluation of adherence was restricted to the availability of handwashing materials during household visits, as they did not perform systematic observations of actual handwashing behavior.

All in all, the promoters made approximately six visits per month on average during the entire trial period, while in the Kenyan study, promoters made an average of one visit every 1-2 months. Furthermore, in both locations, the frequency of the household visits was high in the first year but slowly decreased in the second. In Kenya, the prevalence of ARI's in the soap provision intervention arm was higher than that in Bangladesh in both years. Also in Kenya, in the first year of the trial, the study reported a significant percentage (77%) of households in the handwashing groups had access to water and soap at a handwashing station. However, in the second year, only a small fraction (21%) had access to water and soap.

The findings may highlight a behavioral pattern of a decrease in adherence to the intervention over time by participants. They may also underscore the potential positive contribution of promoting behavior change in a rigorous and persistent manner to achieve meaningful reductions in respiratory infections that have clinical significance.

Lastly, a recurrent habit that was that noted in all trials investigating the effectiveness of water quality improvement on health outcomes was simultaneous consumption of untreated water. This observation was consistently expressed across all of the trials as a limitation to the effectiveness of the intervention and to the evaluation of its impact since its effectiveness could be masked by other recontamination pathways. One trial, (Rosa

et al., 2014), even went as far as stating that “even occasional consumption of untreated water can vitiate the health benefits associated with improved water quality interventions”.

4.4.3 TECHNICAL OUTCOMES

In general, all of the interventions implemented in the trials of the dataset were technically effective.

A. POU HOUSEHOLD WATER TREATMENT

Water Treatment

For water treatment hardware, the following findings were extracted for each intervention type.

1. Chlorination: Chlorination was widely used as a hardware water treatment intervention and it was the most effective microbiologically with bacterial reduction ranging between 86 and 95%. In certain studies, (Albert et al., 2010; Solomon et al., 2020), excessively high residual chlorine was detected, not enough to be health affecting but high enough that they may be concerning with persistent use. Whenever used on water storage, it provided residual protection against recontamination but the percentage was lower in most of the tested stored water, with a reduction ranging between 55 and 72%. As for turbidity, the effectiveness of chlorination in treating it was not tested in the trials that had it as a standalone treatment.

2. Filters: Filters were the second most effective treatment for bacterial contamination. For the reduction in turbidity, the filters were effective but were easily clogged and in the cases where the water was highly turbid, such as in (du Preez et al., 2008; Fagerli et al., 2020), the filters needed to be replaced. That is a problematic observation as in all of the studies that featured filters, the affordability and maintenance of the device were always listed as potential drawbacks, especially in the rural areas settings.

For both bacterial contamination and turbidity, the different types of filters were ranked as follows in order of effectiveness: hollow fiber filters followed by ceramic filters then biosand filters. However, (Stauber et al., 2012), justified their choice of biosand filters by reporting them to be cheapest of all of the other types of filters.

One unique variation of the filters focused upon in the studies in the dataset was the colloidal silver ceramic filter, investigated in (Morris et a., 2018). The trial assessed the performance of silver colloidal ceramic filters in improving *E. Coli* and *Cryptosporidium* rates in the water and reducing the prevalence of diarrhea. The household water samples were only tested for *E. Coli* though and children stool was tested for *Cryptosporidium.* The study found that while the filters were successful in reducing *E. Coli* contamination for 70% of the samples, *Cryptosporidium* prevalence in children stool compared at the beginning at end of the trial experienced a clinically insignificant reduction. The filters were effective in reducing water turbidity, but data about the initial turbidity and its improvement were not reported.

Finally, during the study period, 10% of the filters needed replacement due to either clogging or breakage. The flow rate was also perceived to have slightly decreased, a general observation that was made without the reporting of any data or concrete deduction due to the short trial duration (4.5 months).

All in all, the water treatment apparatus that needed repairing or replacing the most was the filter. In all of the studies that filters, a considerable amount of them were reported to have needed repairing (13-38%) or replacement (9-35%). Furthermore, in (Boisson et al., 2009; Morris et al., 2018; Fagerli et al., 2020), the generated flow rate of the filters was observed to be decreasing over time.

Lastly, one recurrent drawback mentioned for filters is their inability to provide residual protection for stored water, and so for maximum safety, households would need to add to their treatment regimen the addition of a supplementary treatment, mainly chlorination.

3. Flocculent-disinfectant: Flocculent-disinfectant treatment in both featured studies was very effective in reducing turbidity, with an efficiency ranging between 81-93%. For bacterial removal however, and when compared to chlorination and ceramic filters, it was found the least effective, but still being able to provide an "acceptable water safety level", a statement made by (Albert et al., 2010), who tested user preference and the technical performance of the three aforementioned treatment hardware. As for its residual protection against recontamination, it was not tested in any of the studies.

4. Solar disinfection (SODIS): In the only study that focused on solar disinfection, its effectiveness in bacterial contamination reduction was the only parameter tested for and was found to be at an average of 64%. The study mentions that although SODIS was found in many other studies to be much more microbiologically effective, the results that they observed were likely due to the fact that the bottles were exposed to less than the recommended 6 hours of sunlight.

For stored water however, it was only found to be 32% effective, as a significant proportion of the households (32%) were found to regularly exceed the recommendation provided at the beginning of the trial of storing the water for no more than 48 hours from fear of recontamination.

Furthermore, a drawback expressed by the study about the feasibility of SODIS is it not being feasible all year round. The participants also expressed the inability to disinfect the required amount of water for the residents of the house and having to wait a long time before consuming it.

WATER SUPPLY AND STORAGE

1. Drinking Water Supply: In (Cha et al.,2015), investigating the effect of improved household water supply through the drilling of boreholes, no indication was made about the technical performance of the boreholes or any mentioning about field challenges encountered while drilling the boreholes or initiating their operation. Likewise, in (Patel et al., 2012) exploring the effects of implementing a hygiene curriculum in schools along with the installation of water points on the premises and the provision of soap at the stations on student's hygiene practices and diarrheal reduction. The study also did not include any observations about field challenges that the implementers went through or how well the water stations were operating or maintained over time.

2. Storage Vessels: (Stauber et al., 2012) was the only study the studied the sole effect of improved storage and transport in villages with relatively good water quality source for 15 weeks in rural Ghana. They found that the improved storage vessels improved bacterial contamination by around 60% for water stored no longer than 24 hours

in the containers. However, that rate decreased over time to reach 43% for water stored for 3 days and 13% for water stored for 5 days.

However, and as is expected, the authors noted that compared to water treatment methods, enhanced water storage options require an upgraded water supply as a fundamental requirement to achieve any beneficial effects on water quality and health. This implies that while previous water treatment approaches could substitute for an improved water supply, upgraded water storage and transportation methods only serve as a complementary addition to an upgraded water supply, which could be viewed as a limitation of these technologies.

SANITATION INTERVENTIONS

1. (Garrett et al., 2008): The study investigating the effects of latrine construction, provision of chlorine bottles along with improved storage vessels, did conclude that the chlorine and storage vessels ameliorate the households water quality but did not include any additional data on the values or evaluation parameters utilized. As for the performance of the latrines, the study mentioned that during the household visits, the state of latrines was recorded in order to monitor latrine performance and maintenance (breakage, cleanliness, maintenance of the lids) but no results were reported concerning these parameters.

2. SHINE and WASH Benefits Kenya: These studies did not report the states of the constructed latrines during household visits, except for signs of usage, to be able to infer the degree of robustness they exhibited (in terms of breakage or clogging).

3. (Quattrochi, et al., 2021), VEA Trial: In VEA, the approach adopted was community-based whereby communities were provided were only provided funding and general counseling to support the action plan devised by the communities themselves to improve their WASH needs. Accordingly, as a measure of intervention success, participants were surveyed and asked to indicate about the states of their water sources and sanitation facilities. Water sources were categorized as improved or unimproved, depending on whether the water is piped into a dwelling, piped into a plot, sourced from a piped/public tap, tube well or borehole, or a protected spring. Improved sanitation referred to the use of ventilation-improved pit latrines, pit latrines with a slab or composting latrine, or a system that flushes/pours waste into a piped sewer, septic tank or pit latrine, as the primary location for defecation. When it comes to water usage, the outcomes were calculated by combining data from all household members for both the quantity and time variables.

Access to improved water sources was reported to have increased by nearly twofold, from 43% to 83%, while access to improved sanitation facilities more than doubled, from 18% to 46%. Survey respondents also reported notable increases in adopting healthy habits, particularly in terms of handwashing and sanitation-related behavior, as well as better management of WASH facilities at the local level and greater satisfaction with and access to water. However, despite the program's emphasis on reducing the burden of water collection for women and children and the connection between WASH-related childhood diseases, mainly diarrhea, and school absenteeism, there was no significant impact on these outcomes.

5
Analyzing the Use of Randomized Controlled Trials

In recognition of the dominance of the RCT design in the dataset, the following chapter aims to further explore the different constituent aspects of the trials and to situate them within the external literature discourse.

5.1 TRIAL DESIGN

A. TRIAL LOCATION AND SETTING

DATASET TRIAL LOCATION AND SETTING

When it comes to study locations, only 9 out of the 49 countries in SSA were represented in the dataset, with Kenya amassing the most trials. In some cases, the settings in which the trials occurred were in many cases pre-set conditions that represented severe baseline parameters, mainly exceptionally bad infrastructure coverage, extremely high rate of child mortality and diarrheal incidence or high level of fecal contamination in source waters.

Furthermore, another observed phenomenon was that most of the trials were done in similar context settings, in that community households were the main targets. Only one study, (Patel et al., 2012), implemented an intervention within a school context.

EXTERNAL LITERATURE ON TRIAL LOCATION AND SETTING

There has been growing literature on the importance of conducting foreign aid trials in contexts that diverge from the "traditional" household settings, such as in refugee camps or settlements, health care facilities, women health clinics, schools, …etc. These settings may pose unique challenges and may require different approaches to WASH interventions. They may need, for example, a special attention to limited availability or accessibility of resources, overcrowding (in case of refugee camps or clinics at overcapacity for instance) or cultural and language barriers. While the challenges and resources required, including financial disbursement, may be much more intensive and/or intricate, conducting RCT's in these settings can provide valuable insights into the effectiveness of interventions in these real-world settings and potentially prevent further health-related crises (Allegranzi et al.,

2016). Such crises are not rare events but a serious looming threat that is reinforced by the conditions that these settings usually reside in.

In the case of refugee camps and settlements for example, disease outbreaks are reported to be the number one health-related concern (WHO, 2018). Indeed, many such episodes have taken place. In 2017, for instance, a cholera outbreak occurred in a refugee camp in Maiduguri, Nigeria, which was caused by poor sanitation and hygiene conditions (Reuters, 2017). Similarly, in 2018, there was a *Hepatitis E* outbreak in a refugee camp in South Sudan, which was also attributed to poor water and sanitation facilities (WHO, 2018).

Concurrently, a systematic review by (Ramesh et al., 2015), investigating the effectiveness of WASH interventions on health outcomes in humanitarian crises from literature from the past 30 years. They found only six papers, each done on a different WASH trial, but all studying the reduction of self-reported diarrhea. Five of the trials revolved around POU water treatment interventions with only one including the measure of *E. Coli* concentrations as an indicator of water quality improvement. The sixth studied the effect of sanitation on diarrheal reductions. Furthermore, two of the studies were unblinded RCT's while the rest utilized uncontrolled longitudinal study designs, reported by the authors to be usually deemed to yield less epidemiological evidence and quality. The review highlighted the need for more studies of higher quality in humanitarian crises settings, investigating outcomes other than diarrhea.

B. TRIAL DURATION

DATASET TRIAL DURATION

Firstly, 48% of the trials in the dataset were shorter than 6 months and only 13% of trials were longer than 12 months. That, coupled with the fact that these two trial categories offered opposing results for the same outcomes, namely the prevalence of diarrhea, illustrates how significant and influential trial duration consideration is.

Secondly, several impactful phenomena could occur throughout the trial that could be missed in case the duration was not sufficient. It was observed in (Fagerli et al., 2020) for example, that the microbiological performance of the investigated hollow fiber filters decreased over time and encountered a notable rate of malfunctions. Filter breakage rate over time was also an issue highlighted by (Morris et al., 2018). Similarly, (Swarthout et al., 2020), a subset of WASH Benefits Kenya trial, noted a decrease in user uptake of the ceramic filters they were studying over a six-months trial.

All in all, the sustainability of WASH interventions beyond the study period may be crucial to ensure the long-term impact and viability of these interventions. Short-term studies may not be able to provide insights into the challenges of maintaining behavior change and infrastructure maintenance after the intervention period ends. Such knowledge could be imperative to drive the long-term sustainability of the intervention.

TRIAL DURATION IN EXTERNAL LITERATURE

The importance of a trial spanning a long-enough period of time has been emphasized across literature (Mosler, 2012; Thys et al., 2015; Chandrasekhar et al., 2016; Gnilo et al., 2020). These papers present different consequences of short-termed trials, with the main concern being that the thorough and comprehensive

representation of the effects of the intervention may not be sufficiently captured. The reported concerns are in part demonstrated by the findings and observations made by the studies featured in the dataset.

Some literature particularly advised on having the trial run for a minimum time of a year in order to appropriately assimilate the degree of receiver adherence. They also highlighted the importance of quality of life measurements and how certain outcomes, especially those related to behavioral change or quality of life, typically need some time before improvements appear (Crocker et al., 2017; Tumwebaze et al., 2018; Gnilo et al., 2020).

Finally, (Freeman et al., 2017; Gnilo et al., 2020) reported that large-scale trials of WASH interventions tend to be funded by multiple donors, and recommend that there may be greater opportunities for funders to merge and add-on each other's efforts to enable supplementary follow-ups or complementary studies.

C. TRIAL RECIPIENT

DATASET TRIAL RECIPIENTS

There was a notable pattern in the design of the trials in terms of their intended recipients and the data they sought to measure. Firstly, most trials no matter their intervention category targeted children. Secondly, the data amassed during the baseline data collection step were not thoroughly disaggregated by gender and socio-economic indicators.

That is despite the fact that the few studies that did disaggregate their data found different results in their investigated outcomes. For example, (Mengistie et al., 2013), who were exploring the effectiveness of chlorination on diarrheal decrease in children, found that the children of the women who had attained primary school had less diarrheal incidences than those whose caregivers were illiterate. This demonstrates the importance of considering aiming the intervention towards a different recipient group thereby targeting different root causes and problematic pathways.

Also, (Garrett et al., 2008), who opted for a community mobilization approach, targeted women leader groups along with other entities such as local government officials and prominent social committees in the community. They first started with promoting the project to them to gain their support and amassed volunteers from these groups that were trained to promote their intervention and accompanied good-practices, namely latrine construction and use and the consumption and safe storage of harvested rainwater. The authors reported that community members, especially the women, related and subsequently responded well with the female volunteers as they were trusted figures in the community. By making use of the disaggregated data during baseline data collection, the authors found that women practiced less open defection throughout the intervention and used more and more rain harvested water to feed their children.

TRIAL RECIPIENTS IN EXTERNAL LITERATURE

Focus on Children

Firstly, while children may be some of the most vulnerable to WASH-related illnesses, such as diarrhea and anemia, the recurrent approach observed in the dataset of directly targeting their health status changes may not always be the most effective strategy. In fact, it is possible that by focusing solely on children's health and diarrheal

episodes, trials may miss opportunities to target, for example, caregivers who may be essential elements in improving the conditions of their offspring. In particular, women are usually the primary caregivers for children and the elderly, and their access to WASH services can have cascading effects on the health and well-being of entire households. For example, a study by (Sommer et al., 2015) found that lack of access to WASH facilities in health facilities disproportionately affects women, leading to poor maternal and consequent newborn health outcomes. Likewise, (Pickering et al., 2018), for example, found that interventions that targeted both households and community water sources in rural Bangladesh led to improvements in both caregiver and child health outcomes. The authors argue that this was likely due to the fact that caregivers were better able to care for their children when they themselves were healthier.

Consequently, targeting caregivers could be one strategy that leads to more sustainable and long-lasting improvements in WASH outcomes. Indeed, Caregivers are often responsible for the maintenance and upkeep of WASH facilities, and by empowering them to take ownership of these facilities, interventions can be more sustainable and have a greater impact over time (Gautam et al., 2020). Therefore, focusing evaluations onto caregivers, their habits, behavioral drivers, hygiene practices and the priorities driving their preferences and influencing their uptake of the intervention, may be indirectly evaluating the root drivers of poor practices and poor health. It may also be assessing the degree to which the intervention may be sustainable in the long-term and identifying technical or intervention design elements that can enhance the intervention's success by altering them.

Practicalities Considerations

Secondly, the exclusion of practical considerations relevant to women in WASH interventions and their evaluations, may instigate negative drawbacks, both for the women themselves and for the success of the interventions in general. Women and girls are often disproportionately affected by inadequate access to WASH services, particularly during menstruation and pregnancy, and may face additional challenges in accessing and using facilities due to social norms and gender-based violence. For example, lack of access to menstrual hygiene management resources can lead to missed school days, reduced productivity, and increased risk of infection (Caruso et al., 2017). Furthermore, women's needs change over their life-cycle, hence WASH service provision would be ideal if it were suitable for different points in the reproductive life-cycle, including menarche (e.g., separate toilets for girls at school, promotion of menstrual hygiene management approaches) and maternity (e.g., WASH in health facilities, promotion of hygienic weaning practices). Caruso et al. (2017) define sanitation insecurity as "insufficient and uncertain access to socio-cultural and social environments that respect and respond to the sanitation needs of individuals, and to adequate physical spaces and resources for independently, comfortably, safely, hygienically, and privately urinating, defecating, and managing menses with dignity at any time of day or year as needs arise".

There are several examples of unintended negative consequences that can result from WASH interventions that fail to consider the specific needs and preferences of women. For instance, a study in rural Ethiopia found that the installation of public latrines in areas without adequate lighting and security measures resulted in decreased use of the facilities by women, who were afraid of being attacked or harassed while using them (Biran et al., 2011). Similarly, in rural India, the construction of community latrines without gender-segregated facilities and sufficient lighting were also reported to have led women and girls to avoid the facilities altogether, resulting in continued open defecation (O'Reilly & Louis, 2014).

These examples demonstrate the importance of understanding the specific WASH needs and preferences of women, especially in terms of safety and practicality, to ensure that interventions effectively meet their needs and do not have unintended negative consequences. Without such considerations, interventions may not effectively address the root causes of poor WASH outcomes or may even perpetuate existing inequalities. All in all, this highlights the importance of including diverse perspectives and experiences in the design, implementation, monitoring and evaluation of WASH interventions.

Data Disaggregation

Lastly, in recent years, there has been a growing recognition of the importance of data disaggregation by sex and other social identifiers in the evaluation of development interventions, as it makes it difficult to identify gender and social disparities and develop targeted interventions to address them.

For example, the United Nations Sustainable Development Goals (SDG's) call for disaggregated data to ensure that "no one is left behind". Several organizations, such as the World Bank, UNICEF, and the WHO, have developed guidelines for data disaggregation in their respective sectors (UNICEF, 2014; WBG, 2015). These guidelines emphasize the importance of collecting data on sex and other social identifiers and provide practical recommendations for how to do so effectively.

Despite these efforts, however, it has been reported by external literature that there is still a need for increased efforts to ensure that data disaggregation is incorporated into evaluations of WASH interventions in rural SSA. A study conducted in Malawi, for example, found that only a minority of the studies included in a systematic review of WASH interventions reported sex-disaggregated data (Thomson & Kateta, 2020). This highlights the need for greater attention to be paid to data disaggregation in the evaluation of WASH interventions to ensure that the impacts of these interventions are accurately understood and effectively targeted to those in need.

D. TRIAL MEASURED OUTCOMES

MEASURED OUTCOMES IN THE DATASET

1. Reporting of the Outcomes: Before diving into the nature of the investigated outcomes, the way their indicators were reported is first considered. Most outcomes were self-reported, including diarrheal prevalence indicators and user uptake activities. Some studies though assessed participant adherence to the intervention through visual evidence of use or treatment consumption amount. However, none of the applied methods were deemed as satisfactory, thereby causing the questioning of the findings of the trials.

Indeed, even when attempting at adopting an alternative reporting method, those approaches were unsuccessful in providing constituent data. For instance, two studies by the same research team investigating the effectiveness of solar disinfection (du Preez et al., 2008) and ceramic filters (du Preez et al., 2011) on the decrease of diarrhea episodes, handed participants pictorial diaries for the daily recording of the events but encountered in both studies a large absence of reported data.

Only one trial studied by (Rosa et al., 2014) included a sub-trial attempting to record user uptake via sensors mounted on the handed out filters. That approach, while notable in its uniqueness, was way too expensive to be realistically implemented in low-resources settings where a considerable amount of the trial expenditure would

likely be directed to more logistical practicalities. Furthermore, and despite its costliness, the sensors still encountered some technical difficulties that impeded their functioning leading to a substantial amount of missing data.

2. Socio-economic Considerations: As previously mentioned, the most recurrently investigated outcome by the studies in the dataset is diarrheal prevalence, followed by water quality improvement. Beyond it, very few other public health or socio-economic outcomes were targeted by the trials. Indeed, only one study, (Quattrochi et al., 2021), had as a primary outcome the degree of reduction of the amount of water collected and the time to do so.

Likewise, few studies evaluated the cost-effectiveness of the interventions. One exception would be whenever chlorination was the offered intervention, whereby the provision of this POU treatment would be justified by it being cheap and easily accessible. The other exception was (Morris et al., 2018) who attempted to infer the severity of diarrheal episodes by changes in the number of healthcare visits, and by doing so exploring changes in household healthcare expenses which ultimately might have some impact on the household income.

Indeed, the lack of consideration of the cost-effectiveness of the intervention within its intended study area may lead to the wasting of limited resources on interventions that might not survive beyond the trial duration. For example, in (du Preez et al., 2008), after conducting a trial that evaluated the performance of England-manufactured ceramic filters for six months in Zimbabwe, reported only at the end of the study that the filter, including its spare parts, was too expensive for such an under-resourced area. That observation was not deducted from the trial itself but made as a general statement at the end of the study. This consideration might have been better made at the beginning of the trial or while deciding on its location.

That being said, most of the interventions applied in the dataset were low cost household POU water treatments. This may suggest that there is an implicit understanding of the financial, logistical and institutional constraints the communities targeted in this dataset face. In order to investigate more these potential constraints, external literature was consulted:

According to (Howard & Bartram, 2003), such areas may be subjected to political volatility whereby the POS water could become compromised at short notice. These areas may be remote and may not be easily accessible, and they may lack the necessary infrastructure that would facilitate POS water maintenance at community level, which may drive up the financial resources required. Furthermore, political instability and conflicts can disrupt the supply chain of WASH-related goods and services, making it difficult to maintain the necessary equipment and supplies. There may also be a lack of institutional support and capacity at the local level to effectively manage and maintain WASH infrastructure, leading to further challenges in ensuring access to safe water and sanitation. These constraints can make it challenging to implement and sustain large-scale interventions, highlighting a possible motive of the trials for choosing low-cost and household-level solutions for WASH in rural SSA. Indeed, such decentralized approach to safe water supply may afford households the ability to govern the state of their water independently from public sources that may become compromised or inaccessible, especially in times of political instability or conflict. By providing low-cost, simple and easy-to-use POU water treatment options, households can take ownership of their water quality and contribute to overall improvements in water safety and health outcomes in their communities.

3. Practicalities Considerations: Firstly, very few studies assessed of the ease with which the end-user would be effectively able and actually willing to incorporate the intervention in their daily routine. Indeed, an intervention need not be only cost-effective and socio-culturally acceptable, but if an intervention is impractical and requires considerable effort to implement it, its uptake by the receivers would not be considerable.

The importance of this parameter could be best demonstrated by the trial conducted in (Boisson et al., 2009), which was one of the few studies to record feedback on user experience and encountered problems. The trial investigated the effectiveness of the portable version of the LifeStraw hollow-fiber filter. This particular study reported the lowest usage rate in the dataset amounting to only 13% of the recipients, and that is despite the filter being handed for free and the recurrent household visits to record diarrheal prevalence during which renewal messaging about the importance of water safety is done. This low level of uptake was found to be due to the high degree of impracticality of using the filter, with respondents reporting back on rather simple unmet considerations that rendered the filter unusable. Evaluating how user-friendly and practical an intervention is and the experience along with the feedback of the participants the can therefore be imperative to assess how likely an intervention would be accepted and sustainably used.

Secondly, very few of the studies reported on the practical challenges they encountered in the field while they were implementing the intervention. Concurrently, none of the studies included any indication about financial issues or unforeseen costs that arose while conducting the trials. Such considerations however, could be quite significant for the success of an intervention. The pilot study of the WASH Benefits Kenya Trial for example, faced considerable challenges in the rehabilitation of existing latrines and collaboration with local contractors and so the trial altered its original strategy to only construct new latrines in case they were absent at the household. Failure to report such incidences prevents future interventions from taking into account those factors and either explore the reasons behind them, find innovative strategies to deal with them or simply to avoid them. as well as for informing future endeavors of such potential obstacles.

Furthermore, considerations about operating and maintaining the intervention beyond the trial's timespan were sparsely investigated. An illustrative example from the dataset is the LifeStraw hollow fiber filter. The filter had four different trials in the dataset (Boisson et al., 2009; Boisson et al., 2010; Rosa et al., 2014; Fagerli et al., 2020) revolving around it, performed in four different countries: Ethiopia, DRC, Rwanda and Kenya respectively. The technology itself got upgraded to a newer model as well as into a portable device version, studies by (Boisson et al., 2009). However, in all three studies of the household filter version, the frequency of breakage was reported to be high. Yet, none of the trials evaluated the feasibility, practicality and time needed to replace or repair a filter in case of breakage or malfunction. This is especially important since the filter is manufactured in Switzerland and distributed in recipient countries by local partnering NGO's. An added consideration would be the remoteness of the studied areas and their potentially inadequate infrastructure.

On the other hand, the only two trials in the dataset, (Garrett et al., 2008; Quattrochi et al., 2021), with approach principles based on community mobilization, opted for interventions that situate the intervention within "real-world settings". They did so by using strategies other than the quite recurrent one of handing participants free hardware with (or without) messaging. Those two trials started by delivering promotional messaging and then either enabled the community to define and solve their self-defined problems in the VEA program or provided subsidies and help in the construction of latrines in (Garrett et al., 2008). Such approaches may be effective in increasing the

likelihood of sustained usage, especially that they promoted ownership of the intervention instead of handing it for free and attempted at providing the necessary knowledge to maintain it.

4. Alternative Contamination Pathways Considerations: The issue of contamination along other pathways, while acknowledged by almost all studies that targeted health-related outcomes as an issue and a possible reason for the non-effectiveness of the intervention, it was not targeted or investigated by any of the trials, except secondarily by those focusing on health and hygiene messaging and those including storage considerations and hardware in their intervention.

LifeStraw: The technology itself got upgraded to a newer model as well as into a portable device version. In the initial trial with the LifeStraw 1.0 by (Boisson et al., 2010) in the DRC, the effluent water quality was found to have been improved but the uptake was low and the filter's effect on diarrheal prevalence was null. While the lack of influence on diarrhea was attributed to the low uptake, the authors emphasized more the high probability of recontamination, especially that the device does not afford residual protection.

Furthermore, although the issue of recontamination was highlighted in all trials, including the study by (Boisson et al., 2009) on the portable version of the filter, the updated version of the filter and the succeeding trials never addressed or assessed the seriousness and influence of the issue except for acknowledging the possibility of recontamination as one of the trial limitations.

Finally, as previously mentioned, *E. Coli*, and less frequently Total Coliforms and turbidity were the most used indicators to assess the degree of water quality improvement and link it to health-related outcomes. However, as mentioned in (Morris et al., 2018) studying the effects of colloidal silver ceramic filters on the reduction of *Cryptosporidiosis*, a severe diarrhea-inducing disease that is chlorine resistant, *Cryptosporidium* is one of the leading causes of diarrhea morbidity and mortality in LMIC's. While this protozoon is harder to measure than *E. Coli*, orienting interventions towards its detection and reduction may instigate a substantive reduction in diarrheal prevalence.

5. Causality Behind the Outcomes Considerations: The causality behind the perceived outcome results was not always sufficiently established in some of the studies in the dataset.

Firstly, studies evaluating the effectiveness of POU household treatments on the reduction of diarrheal episodes started with the hypothesis that the improvement of water quality leads to a reduction in diarrhea. However, some of these studies have failed to measure changes in water quality due to the intervention. This is especially problematic since POU treatments were reported in some studies to have experienced technical challenges and breakage, such as in the case of filters, thereby compromising their water treatment efficiency. As a result, without considering water quality, it becomes difficult to establish a causal relationship between the intervention and any potential reduction in illness episodes. The lack of measurement of water quality before and after the POU treatment intervention can therefore lead to an inaccurate assessment of the effectiveness of POU treatments.

Secondly, and as previously mentioned, the reasons behind the results obtained were not investigated in many of the studies in the dataset. This can make it difficult to deduce any causal impacts of the intervention. For example, a study may find that the installation of a latrine or a POU treatment led to a reduction in diarrheal

episodes. However, without assessing the reasons for this reduction, it would be difficult to deduce if it is due to the intervention itself or other factors, such as increased handwashing or improved water quality. Additionally, studies often overlooked the importance of evaluating why the uptake of certain interventions was successful or not, including factors such as seasonality and its impact on changes in water quality and turbidity, cultural practices, or other parameters. The issue is exacerbated by if a study has a short duration, which may limit the ability to observe or predict long-term effects and understand the complex social, cultural, and environmental factors that influence the effectiveness of WASH treatments and their uptake. Such information may be critical for designing effective and sustainable WASH programs or interventions.

Finally, certain studies did not measure the necessary data during the baseline data collection method, which compromised the ability establish a solid causal relationship between the intervention and its outcomes. Without sufficient and diverse baseline data, the differences observed between intervention and control groups may not be attributable to the intervention, thereby creating a risk of bias in the results.

TRIAL MEASURED OUTCOMES IN GENERAL LITERATURE

Firstly, (Scott et al., 2007) argued that evaluating the efficiency and sustainability of a technically-effective intervention requires the exploring of several critical parameters that should mainly ensure two overarching characteristics: recipient-uptake and the long-term sustainability of the intervention. According to (Waddington et al., 2009), in order to ascertain these two parameters, an intervention must possess three main elements: it should provide recipients with access to a functional intervention and enough knowledge to know how to operate it and why it is important to do so, and, most importantly, that knowledge ought to be practiced. Indeed, all of these parameters are important determinants for the uptake of an intervention and the duration of its employ (Prüss-Üstün et al., 2019).

Likewise, (Biran et al., 2012) states that while the principles behind an intervention and its outcomes can be straightforward especially when it comes to its technical functionality and operational soundness, the intervention operates within a socio-economic and cultural frame and is geared by behavioral dynamics, both of which strongly determine its acceptance, application and effects in real-life settings.

Behavioral dynamics are guided by the beliefs, experiences and values of recipients within the social structure. Therefore, elements like the social connections a recipient has with their community as well as individual agency are imperative in the uptake or dismissal of an intervention. The authors further argue that behavioral change is attained through the context, which they define as the “entire system of social relationships”, and therefore assert that an intervention or program directed by or towards behavioral change ought to line up with the context in which it is employed.

Indeed, the consideration of behavioral dynamics in the design and implementation of a trial, including the user practicality of the intervention, is especially since different interventions require different rates of behavioral change and adaptation. For example, water supply and the implementation of sanitation hardware do need some behavioral alteration to insure that they are used and well-maintained, especially hygienically to limit the contraction and spear of microorganisms (WHO, 2019). On the other hand, interventions that target water quality or hygiene, require considerable and rigorous behavioral alteration, especially with interventions that require recurrent, systematic and

sometimes time-intensive or laborious application; or in the case of hygiene promotion, the need for handwashing repeatedly all along the day for example (Fisher et al., 2018).

Secondly, and on the same note, when it comes to exploring the causality between the outcome and the intervention, (Waddington et al., 2009) argued that an assessment ought to investigate beyond the determinants of technical, health and straightforward behavioral indicators and attempt at understanding why and the reasons behind an intervention generating the outcomes that it did.

Indeed, links in the causal pathway between interventions and outcomes are not automatic. For example, latrine building may not reduce open defecation or contamination in the public domain. Factors limiting use include the cleanliness and smell of the facilities, or concerns about how frequently the pit will need to be emptied. Latrine provision may also not improve health and nutrition where open defecation is practiced in densely populated areas (Geruso & Spears, 2018; Kar & Chambers, 2008). Children may be afraid of going into dark places or of falling into latrine pits, creating further hazard since young children's excreta contain the most pathogens (Curtis et al., 1995).

E. TRIAL DESIGN

DESIGN OF TRIALS IN THE DATASET

1. Randomization: Randomization was not always done according to "good practices", hence potentially introducing bias. In (Stauber et al., 2012), the randomization was done on a village level instead on household level. While consulting the general literature about the potential consequences of this design practice, two potential effects were recurrently reported across several studies (Bastos Lima et al., 2015; Hayes et al., 2016; Chow et al., 2018). They suggested that scaling up randomization could impact the study's validity and the generalizability of the findings since it assumes homogeneity within villages, implying that all households within a village are similar and have the same characteristics. However, in reality, there can be significant variation in households within a village, such as differences in income, education, and sanitation practices. If randomization is done at the village level, this heterogeneity is not accounted for, and the treatment group and control group may not be comparable, leading to biased results. Furthermore, individuals within a village are assumed to be independent of each other and that the treatment effect will not spread from one household to another. However, it is possible that the treatment effect will diffuse across households within the same village, leading to contamination of the control group and an underestimation of the treatment effect.

2. Inclusion and Exclusion Criteria: When it comes to choosing to exclude a certain portion of the population, bias may be generated. By excluding participants who did not adhere to the intervention protocol and guidelines in (Cha et al., 2015), or severely ill caregivers and children with persistent diarrhea while the intervention was geared towards the reduction of diarrheal prevalence in (Solomon et al., 2020), a certain level of bias may be imposed on the studies' findings, especially that the study did not add any additional information about the number of participants they excluded.

For more information, external academic literature was consulted to explore potential resultant repercussions, and the following was identified: if the intervention is more effective or less effective for severely ill individuals, excluding them can result in an over- or underestimation of the true treatment effect. This is especially true in low-income

areas whereby the chances of disease outbreaks or high fecal contamination rates are substantial, and the proportion of the community affected by them were considerable (Whitehead et al., 2016).

3. Choice of Baseline Conditions: Choosing study areas with exceptionally poor baseline conditions RCT's such as in the cases of (du Preez et al., 2008; Boisson et al., 2010; Rosa et al., 2014; Cha et al., 2015; Fagerli et al., 2020) can introduce bias into the study results in several ways. Since no extra information concerning the potential effects of choosing such conditions were explored within the studies in the dataset, external literature was consulted and the following effects were found:

a. Firstly, if the baseline conditions are extremely poor, there may be a ceiling effect, meaning that there is limited room for improvement in the outcomes being measured. As a result, even if the intervention is effective, the improvement seen in the treatment group may not be statistically significant compared to the control group.
b. Secondly, if the baseline conditions are very bad, it may be more difficult for the control group to show improvement, even without the intervention. This could bias the results in favor of the treatment group, making the intervention appear more effective than it actually is.
c. Finally, the "low hanging fruit effect" further complicates matters by suggesting that the greatest improvements may be seen in areas with the worst baseline conditions, but these improvements may not be sustained over the long term. The low hanging fruit effect can lead to exaggerated effect sizes, which may not be reproducible in other settings with less extreme baseline conditions (Sood et al. 2010).

4. Blinding and Self-reported outcomes: Binding, or the lack thereof, was a recurrent limitation across almost all the trials in the dataset. The fact that almost all of the studies in the dataset were unblinded underscores the complexity of implementing placebo interventions in the field. Likewise, the dominant utilization of participant-reported data across trials, its repercussions on the trial findings and incurred potential of bias have already been detailed several times in this study. These two design choices exacerbate each other and may the rigor and acceptability of a trial.

In the dataset, some trials reported the reason for their choosing not to blind, the common reason was in the case of finding that the POS or POU water at baseline was turbid, and so an ineffective treatment would be easily detected. However, most of the studies did not report on the reasons behind not blinding, except for some citing the failed blinding attempts done in (Boisson et al., 2010) as a motivation not to blind. Similarly, the reasons behind choosing to use self-reported data instead of quantifiable and measurable data was not reported.

For more information on the potential repercussions of not reporting the reasons behind not blinding, external literature was consulted, and the following was identified:

The lack of reporting on the reasons behind not blinding can have a range of potential effects on trials, including reducing transparency, hindering reproducibility, increasing the risk of bias, and reducing the overall quality and rigor of the study. Furthermore, the lack of reporting on the reasons in LMIC settings may have unique implications related to resource limitations, cultural and linguistic barriers, and potential harm to study participants. Failing to report on these challenges and limitations may also compromise the transparency of the trial.

a. Reproducibility of the Findings: The lack of reporting can make it difficult to reproduce the study or apply its findings in other contexts. If researchers do not have a clear understanding of the reasons why blinding was not possible or feasible in a particular trial, they may not be able to replicate the study design, adapt it to other settings, or account for and find innovative methods to bypass the challenges (Savovic et al., 2012). This is especially true in the context of LMIC settings whereby unique logistical, financial, cultural or linguistic challenges and limitations may arise impeding the ability to blind a trail. For example, lack of resources or infrastructure may limit the ability to implement blinding or placebo interventions in clinical trials. Authors may choose not to blind because it may be logistically challenging or costly to implement a placebo intervention in a rural setting with limited resources.
b. Cultural and Linguistic barriers: Cultural and linguistic barriers may also play a role in the implementation of blinding in LMIC settings. For example, community members may be less likely to trust unfamiliar or foreign treatments, or the concept of a placebo intervention may be hard to communicate. In such cases, researchers may choose not to blind to avoid potential harm or confusion among study participants (Peprah et al., 2021). Secondly, the lack of reporting on reasons for not blinding can make it difficult for researchers to identify potential sources of bias or confounding variables that may have influenced the study results (Viergever, & Ghersi, 2011).

TRIAL DESIGN IN EXTERNAL LITERATURE

1. Debate Around RCT's: RCT's are hailed in development research as the gold standard for evaluating the effectiveness of interventions (Glasziou, et al., 2004).

a. Proponents of RCT's : They argue that these trials provide a high level of reliability and unbiased evidence for evaluating the safety and efficacy of interventions. RCT's are contended to minimize the effects of confounding variables by randomly assigning participants to intervention and control groups, allowing for causal inferences to be made about the effects of the intervention. For context, a confounding variable is a variable that is related to both the treatment being studied and the outcome being measured, and can therefore lead to biased or misleading results if not properly accounted for in the study design and analysis. Examples of confounding variables in RCT's could be baseline characteristics of study participants, such as age, gender, socioeconomic status, or health status. If these characteristics are unevenly distributed between the treatment and control groups, they could affect the outcome of the study and make it difficult to determine whether the treatment or the confounding variable is responsible for the observed effect (Djulbegovic & Guyatt, 2017).

   Additionally, RCT's are intended to generate findings that are generalizable aided by its focus on minimizing bias. The generalizability of a trial refers to the extent to which the results of a study can be generalized to populations and settings beyond the study sample and context. It is deemed important since it determines the applicability of the findings to real-world contexts and widen the impact of the intervention. In fact, it is reported as the number one motivation for employing this trial design (Higgins et al., 2017). Conversely, the consequences of compromised generalizability in RCT's can be significant. If an RCT has limited representativeness, this can result in interventions that are not effective or not as effective as they could be in populations with different characteristics or experiences. This can lead to wasted resources, missed opportunities for improving health or social outcomes, and potential harm to participants who receive

ineffective or harmful interventions (Kaptchuk et al., 2018). Additionally, compromised generalizability can undermine the credibility and validity of RCT's as a scientific method for evaluating an intervention, as it can raise questions about the relevance and usefulness of the findings in informing policy and practice.

That being said, although RCT's may not be expected to yield universally applicable results, they should however be constructed and presented in a manner that enables practitioners to assess their applicability to specific recipients, environmental and socio-economic contexts, and/or periods of time. To do so, the standardization of study design and outcome measures was reported to be a critical aspect of research design to ensure that the results are comparable across different settings. Standardization of study design involves ensuring that the study design is consistent across different settings. This includes aspects such as the sampling strategy, data collection procedures, and data analysis techniques. This is important because if, for example, one study uses a cluster randomized design, while another uses an individual randomized design, it may be difficult to compare the results of the two studies as the design differences may have affected the outcomes. Standardization of outcome measures is also important to ensure that the outcomes are comparable across different studies. This involves selecting outcome measures that are valid, reliable, and sensitive to change, and that are appropriate for the study population and intervention. For example, if one study uses self-reported measures of diarrhea, while another uses stool culture testing, it may be difficult to compare the results of the two studies as the different measures may have different levels of sensitivity and specificity (Bhutta et al., 2005).

b. Opponents of RCT's: On the other side of the debate, opponents of RCT's on the other hand have pointed out one recurrent argument across literature. It states that the generalizability the findings of RCT's may be easily compromised by several factors including the introduction of bias in the findings. This concern was one that was repeated extensively across literature, which included multiple aspects that could constrain RCTs' ability to be scaled up and replicated in other settings (Tambo et al., 2014; Georges et al., 2015; Bauer et al., 2016; Parpia et al., 2016; Peletz et al., 2016; Freeman et al., 2017).

The generalizability of an RCT is closely linked to its representativeness, or the degree to which the study sample reflects the characteristics of the broader population, therefore if an RCT is restrictive in its design, the representativeness of the study sample may be limited and bias may be introduced into the study findings. One common trial design characteristic that was reported to have been vulnerable to bias in RCT's is the inclusion and exclusion criteria. Indeed, if for instance, an RCT only includes participants who are highly motivated or have a certain level of education or income, the results may not be generalizable to the broader population who may have different characteristics, priorities or experiences (Higgins et al., 2017). Additionally, the short follow-up periods commonly used in RCT's may not allow researchers to observe the long-term or comprehensive effects of an intervention, which could further limit the generalizability of the findings (Horwitz & Horwitz; 2017). Finally, RCT's are often conducted in highly controlled settings, which may result in outcomes that are not representative of what would occur in real-world settings, which could further overlook contextual factors that may affect the effectiveness of an intervention (Eaton, 2018).

2. Comparability of the Dataset to the General Literature: In the context of the topic of this study, evaluation studies of foreign aid interventions in the water sector of the rural regions of LMIC in general as well as SSA in

particular that were featured outside of the dataset were reported to conduct trials that mimic the approaches and different elements observed within the evaluative studies of the dataset. These are essentially the preference for the RCT design and the difficulty in blinding, the popularity of the traditional low-cost water treatment hardware, the dominance of water treatment-targeted interventions over other fields in WASH development sectors, namely sanitation hardware implementation, hygiene messaging and water supply interventions such as the installation of water points or borehole drilling. Furthermore, the focus on the household scale and settings, as well as centering interventions around children were also other common characteristics. Finally, the discrepancies observed in the results of the trials, mainly those investigating the reduction in children diarrhea prevalence, were also observed in impact evaluations outside of the dataset that were conducted in different LMIC around the globe.

3. Alternative Trial Designs: As a result of these concerns, alternative study designs that are less costly and potentially more feasible to evaluate the impact of WASH interventions in these settings are being recognized more and more in general literature.

1. Before and After Concurrent Control (BAC): These studies are a type of quasi-experimental study design that can be used to evaluate the impact of an intervention. The design involves measuring outcomes before and after the intervention in the same group of participants. Participants serve as their own control, with the before and after measurements serving as a comparison of outcomes.

   BAC studies compensate for the limitation of RCT's in terms of cost and feasibility. In an RCT, both intervention and control groups must be established upfront, and it may be challenging to identify and recruit a suitable control group. BAC studies overcome this limitation by using the same group of people as both intervention and control groups, and by measuring the outcome before and after the intervention. BAC studies can also provide more detailed information on the timing and magnitude of the impact of the intervention since outcomes are measured at two time points.

   However, BAC studies are still subject to some limitations, such as the potential for selection bias and the inability to fully control for confounding factors. Additionally, BAC studies do not provide information on the long-term impact of the intervention, as outcomes are only point measurements made before and after the intervention (Langlois & Miszkurka, 2019).

2. Stepped-wedge Trials: These trials are another alternative study design that can be used to evaluate the impact of WASH interventions. These trials are a type of RCT in which the intervention is introduced in a randomized sequence across different clusters (e.g., communities or health centers) over time, with all clusters ultimately receiving the intervention. In this way, the trial is "stepped" from the control phase to the intervention phase over time. Each cluster serves as its own control during the control phase, which helps to control for potential confounding factors.

   Stepped-wedge trials compensate for the ethical limitation of RCT's, which require withholding the intervention from some participants (more elaboration on RCT ethical considerations are elaborated upon in the next subsection). They ensure that all participants ultimately receive the intervention, while still allowing for a rigorous evaluation of the intervention impact. They can also be more feasible than RCT's in

certain contexts, as the intervention is rolled out gradually over time rather than all at once. Stepped-wedge trials can also provide information on the timing and magnitude of the intervention impact, as outcomes are measured at different time points.

However, stepped-wedge trials are also subject to some limitations, such as the potential for contamination and time-varying confounding factors. Contamination can occur if participants in the intervention clusters interact with those in the control clusters or if there are external factors that affect both clusters. Furthermore, the design may require longer study periods than other designs, as all clusters must serve as controls before receiving the intervention. The design can be more complex to implement than other designs, as the rollout of the intervention must be carefully planned and coordinated. There may also be concerns about participant retention over the long study period. (Hemming, et al., 2015).

5.2 ETHICAL CONSIDERATIONS IN RCT'S

A. ETHICAL CONSIDERATIONS IN THE DATASET

Ethical considerations in the dataset were generally standardized. Written informed consent from the participants was obtained at the beginning of the trials, along with an approval from two independent ethical committees, one belonging to the research institution conducting the trial and one belonging to the affiliated organization in the country in which the trial takes place. As for ethical considerations within the recipient communities, only 34% of the studies mentioned seeking the approval from village leaders or head figures within the area.

Beyond the participant consents and approvals from community leaders, few ethical considerations were reported. Only one study, (Stauber et al., 2012), mentioned that one of the reasons for the lack of introducing a placebo filter was for ethical considerations, essentially since caregivers were reported to filter their water by boiling it when for their child's consumption. The authors hence hypothesized that the ineffective filter would be used for child consumption and expose the recipient to potential harm.

B. ETHICAL CONSIDERATIONS IN EXTERNAL LITERATURE

While browsing through the external literature of ethical considerations to be taken for RCT's, especially those performed in LMIC settings, the following points were made by the general literature.

POWER IMBALANCE

Firstly, foreign aid WASH RCT's in low-resource settings can be impacted by a complex web of ethical concerns, including power imbalances, accountability, and foreign intervention. The legacies of imperialism and colonialism play a significant role in shaping power dynamics and resource distribution between the Global North and the Global South, potentially perpetuating unequal exchange and power imbalance.

Power imbalances can manifest in a number of ways in these studies, including potentially leading to the instrumentalization of the participants. Instrumentalization refers to treating individuals or communities as means to an end rather than respecting their dignity and autonomy. In the context of RCT's, this means using participants solely for the purpose of conducting research, without considering their needs and interests beyond the scope of

the study. This can be problematic as it can potentially lead to exploitation. Exploitation refers to the taking advantage of vulnerable individuals or communities who may lack access to healthcare, education, or other resources. This may lead to participants not understanding the nature and purpose of the study, thereby affecting the validity of their consent (Sultana, 2018).

Additionally, researchers may approach the community with preconceived notions about what is best for them, which may lead to a lack of community input in the study design and implementation process. These issues can create a situation where the local community is not able to fully express their own needs or preferences, and the choice of who receives treatment may not be made in a manner that reflects local priorities and concerns (Ruppel et al., 2019).

Likewise, another consequence of these legacies is the potential for researchers to approach the community with a patronizing or paternalistic attitude, viewing themselves as the experts with solutions to offer. This can lead to a situation where the community's knowledge and expertise is undervalued or ignored, potentially contributing to a lack of sustainability or cultural appropriateness in the intervention. Additionally, the history of exploitation and extraction in these regions can create a situation where the community is wary of outsiders and skeptical of their intentions, potentially leading to a lack of trust or cooperation between the researchers and the community.

The legacy of imperialism and colonialism can also contribute to unequal exchange, where the benefits of the study are not distributed equitably between the Global North and the Global South. This can manifest in a number of ways, such as researchers from the Global North receiving more credit or recognition for the study than their local collaborators, or the benefits of the study being primarily directed towards the interests of the researchers or funders from the Global North rather than the local community.

Furthermore, the colonial dimensions of Global North-based researchers intervening to estimate what is best for people in the Global South can contribute to a situation where the study perpetuates rather than challenges colonial power dynamics and the values and beliefs of the Global North stakeholders may be imposed without proper consideration for the local context. This can be problematic as it may lead to interventions that are not culturally appropriate or sustainable in the long term. This can also lead to a situation where the local community is not seen as an equal partner in the study, but rather as a passive recipient of interventions and knowledge (Robins, 2018).

The consent process can also be impacted by power imbalances, as individuals may feel pressure to participate in the study due to external factors or unequal power dynamics. This can lead to a situation where consent is not truly informed, and participants may not fully understand the risks and benefits of participating. This can be exacerbated by the fact that obtaining informed consent can be challenging in low-resource settings, where participants may have limited access to information or may not fully understand the research process. Additionally, cultural differences and language barriers can affect the consent process, further highlighting the need for clear and effective communication between researchers and participants.

Finally, power imbalances can impact accountability, as researchers may prioritize the interests of their funders or other external stakeholders over the interests of the local community, potentially overlooking the effects of the intervention on the participants (Harcourt, 2012).

EQUIPOISE

The second ethical consideration that was noted in external literature was the notion of equipoise and the importance of its presence. Equipoise refers to the state of genuine uncertainty regarding the comparative effectiveness the treatments or interventions being studied in a trial (Freedman, 1987). Equipoise is essential to maintain ethical justification for randomizing participants to different treatment arms, particularly in the context of LMIC settings such as rural SSA. These trials are typically conducted in low-resource settings where the population is often vulnerable and in need of better outcomes (Gill et al., 2013). In this context, equipoise requires that researchers are uncertain whether the intervention being studied is effective or not, and that they are willing to accept the possibility and risk that it may not be effective (Freedman, 1987). This is important to avoid exploiting vulnerable populations and to ensure that limited resources are allocated in an optimal way and that the trial results will inform future clinical practice and policy decisions (Miller & Joffe, 2016).

However, achieving equipoise can be challenging due to several factors, including pressure from funding agencies and donors to demonstrate success (Gill et al., 2013). To address these challenges, researchers are recommended to carefully consider the design of the trial, ensure that the intervention is evidence-based and culturally appropriate, and engage with the local community to ensure that the trial is conducted ethically and transparently. By doing so, researchers can help to ensure that the trial maintain equipoise and contribute to improving health outcomes for vulnerable populations (Torgerson & Torgerson, 2010).

WITHOLDING TREATMENT

Thirdly, the use of RCT's poses ethical questions as they require the random selection of participants who will receive the intervention, with those in the control group receiving none or an inferior one. This can be objected to on ethical grounds, such as Kant's belief that it is always wrong to use human beings as a means to an end, with critics arguing that RCT's prioritize learning over the well-being of study participants. However, completely rejecting RCT's on this basis would also prevent the development of new innovative interventions, the determinations of the efficiency of interventions and the assurance that resources are being allocated optimally (Emmanuel, 2000). Furthermore, (Singer et al., 2019) have argued that prohibiting people from participating in such trials would be overly paternalistic and infringe on their personal freedom and they concluded that while RCT's are not without ethical challenges, they can be conducted in an ethical manner with appropriate safeguards in place.

5.3 TIME AND FINANCIAL REQUIREMENTS OF RCT'S

One chief consideration that was poorly reported and focused on in the papers featured in the dataset of this study is the financial and temporal requirements of conducting RCT's and their effects on the successful implementation of the trial and the drawn findings.

This is despite the fact that while exploring external literature, several sources attributed some of the potential shortcomings that were discussed in the above to RCT's being expensive and time-consuming, thereby hindering the optimal adaptation of such trials in resource-limited settings. This is explored in detail in the following:

1. The time and financial requirements of WASH RCT's in rural SSA may pose significant challenges for both researchers and funding agencies. Such trials can require extensive investments in planning, implementation, and monitoring, with limited room for error, given the high costs of conducting it.

2. The time and financial constraints could also lead to difficulties in recruiting sufficient participants and retaining them for the duration of the study, especially that RCT's usually require large sample sizes to achieve statistical power. This could also lead to difficulties in ensuring adherence to the intervention and control protocols, through for example, the limiting of the amount, frequency or quality of the intervention (Peletz et al., 2016). Concurrently, ensuring the ethical conduct of the study can also be costly, including obtaining informed consent, ensuring participant safety and understanding of the trail, its general principles and mechanisms, and protecting sensitive data (Cope et al., 2018).

3. In addition, the need for specialized equipment and expertise, as well as infrastructure and supply chain challenges, can further add to the costs and complexity of conducting WASH RCT's. This is especially true in low-resource settings and remote areas such as the ones considered in this study, whereby logistical challenges such as transportation, infrastructure, and staffing, are substantially more likely to arise and could further increase the cost and time required for the study (Freeman et al., 2017).

4. Furthermore, the time and financial requirements may also lead to a focus on short-term outcomes, rather than long-term sustainability. Some interventions may also require longer follow-up periods and greater investment in data collection, especially in order to obtain a comprehensive overview of their effects, which could increase the cost of the study and the time required to complete them. All of the aforementioned could constrain the potential impact of the intervention (Tambo et al., 2014).

5. Finally, the costs associated with conducting RCT's in low-resource settings may limit the ability of local governments and organizations to sustain interventions beyond the study period, further limiting the potential impact of the intervention (Georges et al., 2015).

6
Discussion

In this study, the major themes and trends of impact evaluations of WASH interventions in rural SSA were identified and extracted from studies featured in the largest development database of systematic reviews and impact evaluations in the world, the Development Evidence Portal (DEP) (3ie Development Portal, n. d.). the study starts with identifying relevant publications in the 3ie DEP pertaining to the water sector in rural SSA based on systematic inclusion criteria. The study goes on to extract the characteristics data of the amassed literature, and perform bibliometric and network analyses for an insight into the research community of the dataset. The study continues by analyzing the results from the characterization, the bibliometric and network analyses and the content of the publications, all the while mirroring those findings with the general literature to identify gaps within the field and finally present some recommendations for future research.

6.1 BIBLIOMETRIC ANALYSIS

6.1.1 TEMPORAL DEVELOPMENT OF THE DATASET

A. Observed Trends in the Genral Water Sector in Globally and in SSA

Firstly, before focusing on the data from the refined dataset that was restricted to WASH-related interventions, the temporal distribution of the number of publications pertaining to observed trends in the general water sector globally and in SSA is considered. Both trends showed an increase in publication rates from 2008.

One possible explanation for the increase in the number of publications from 2008 onwards is the International Year of Sanitation appointed by the United Nations General Assembly in 2008, with the aim of increasing awareness about the importance of sanitation for human health and well-being. The choice of the year 2008 as a milestone was significant since it represented the midpoint towards achieving the Millennium Development Goals (MDG's) by the target date of 2015, which aimed to reduce the proportion of individuals without access to basic sanitation by half. Several activities were carried out at the local, national, and international levels throughout the year to increase awareness about sanitation issues, including community-based projects, advocacy initiatives, and public education campaigns (UN, 2008). This campaign helped to draw global attention towards the need to invest in sanitation infrastructure and services for promoting sustainable development and reducing poverty. The campaign also successfully attracted significant donors, including the Bill and Melinda Gates Foundation, towards a sector that was

earlier believed to be too costly for effective impact evaluations to be conducted (UN, 2014). Seeing as the Gates Foundation was the funder of 35% of the studies in the dataset, along with a number of other foundations, international agencies, international research institutes and research universities, these contributors could be the catalyzers in impact evaluations publications after 2008.

Furthermore, an increase in the publication rate for both datasets was also observed from the year 2018. The launch of the United Nations Decade of Action on Water in 2018 may have contributed to this increase. The Decade of Action on Water aimed to ensure sustainable management of water resources and sanitation for all by 2030. This initiative is a global effort to address water-related challenges and promote sustainable water management practices. The launch of the Decade of Action on Water may have served as a catalyst for increased research and publications on water-related development issues, as it draws attention to the importance of water resources and the need for sustainable management (UN Water, n.d.).

B. Observed Trends in the WASH Sector in SSA

Secondly, zooming in on WASH-related interventions specific to SSA, the dataset exhibited an increase in yearly publications after 2008. The recurrence of the two main types of interventions before and after 2009 is depicted in Figure 15. These interventions were found to be disproportionally related to household POU water treatment hardware (with or without health and hygiene messaging), followed by those focusing on safe sanitation opting for the decrease in open defecation.

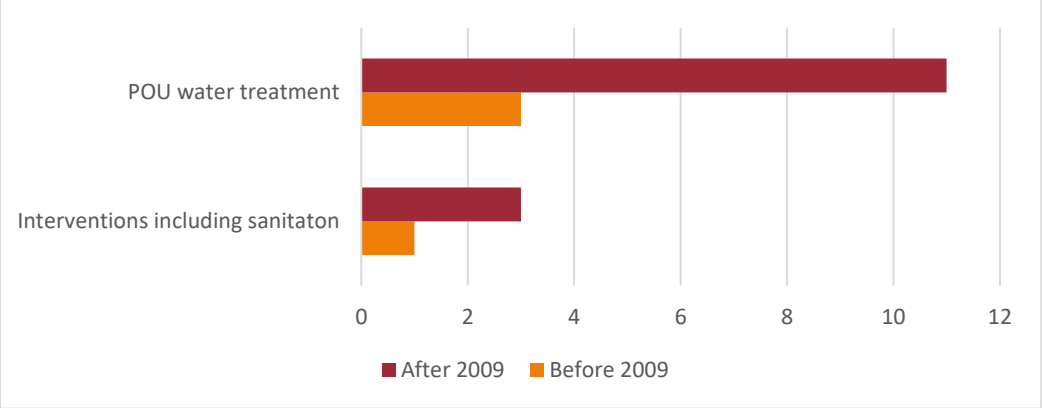

Figure 15: Recurrence of the Two Main types of Interventions Before and After 2009

Since the trials that include safe sanitation in their interventions were found to be mostly in large-scale trials, then the higher rate of recurrence of POU treatment interventions over sanitation interventions may be expected. However, that being said, there has been an increase in the number of sanitation interventions from 2009 onwards. Furthermore, seeing as there was an increase in the number of publications related to international development in the water sector in general from 2008 and that this trend was translated in both dataset related to global trials and those in SSA, a spillover effect might be hypothesized that would have led to more recognition of the field of

development in the water sector in general and which would also explain the increase in the number of publications in all both dataset and in both types of interventions.

6.1.2 PUBLISHING JOURNALS

A. COUNTRY OF ORIGIN OF THE JOURNALS

The concentration of journal publications from high-income countries like the UK and USA in the WASH interventions in rural SSA dataset is not surprising, given the higher level of research funding and capacity in these countries (Tripathy et al., 2016).

However, the lack of publications from developing countries in the dataset is concerning, as it may suggest a limited capacity for research and evaluation of interventions in these settings. This may highlight the need for capacity-building efforts and research collaborations between high-income and low-income countries to promote more equitable and comprehensive research on WASH interventions in rural SSA. This is important as research conducted in developing countries may provide unique perspectives and insights into the effectiveness of WASH interventions in different cultural and environmental contexts, which can inform the development of more context-relevant interventions (Ware, & Mabe, 2015).

B. RESEARCH FIELDS OF THE JOURNALS

Approximately half of journals (48%) in which the WASH studies were published are related to tropical medicine and health research, specifically in the areas of epidemiology, infections and diseases. This suggested that there may be a strong focus by the studies on addressing health-related issues and improving healthcare outcomes through WASH interventions. This is in line with the broader global health agenda, which is increasingly recognizing the importance of WASH interventions in promoting health and well-being, particularly in LMIC's (Bain et al., 2014).

Furthermore, almost a third of the journals are related to environmental science, which may mirror a prevalent recognition in literature of the importance of the environment in sustaining health and well-being. This could be in line with the understanding that WASH interventions are closely linked to environmental conditions and that improving access to clean water and sanitation can have a positive impact on both environmental and health outcomes (WHO, 2019).

The relatively low portion of journals related to economics (16%) and development (7%) in the dataset of publications around WASH RCTs in rural SSA suggests that research on WASH interventions in this context is still largely driven by the public health perspective, with a focus on improving health outcomes rather than economic development. This may reflect the priorities of the funders and stakeholders involved in the projects included in the dataset.

However, these approaches are in contrast with the growing recognition of the complexities and requirements of evaluations of WASH development interventions. This can be especially true in resource-limited settings where funding for research, adequate infrastructure, and trained personnel may be limited (Biran et al., 2012). Indeed, economic considerations have been recognized as foundational considerations when planning and conducting WASH intervention evaluations, as they can impact the feasibility and rigorousness of the evaluations, which may also

eventually affect the success and sustainability of these interventions in the long term. Well conducted cost-effectiveness analyses, for example, can help to determine the most efficient allocation of resources for WASH interventions, while economic evaluations can provide valuable information on the return on investment of these interventions (Dreibelbis et al., 2013).

Innovative study designs and data collection methods can help to ensure that evaluations of WASH interventions provide cost-effective and informative results. Although there has been new research on innovative study designs and data collection methods that can be used to overcome the aforementioned challenges, these approaches are still underdeveloped, impractical and costly (Kazemi et al., 2019). Furthermore, the use of these methods has not been widely recognized yet or markedly featured in economic journals (Peprah et al., 2021). This may reflect a gap in the recognition of the importance of incorporating and improving economic considerations in WASH interventions and evaluations.

Likewise, the low number of publications published in development journals contrasts the growing literature on the potential economic and social benefits of WASH interventions in promoting economic development, particularly in rural areas of low-income countries. Indeed, the World Bank has highlighted the potential for WASH interventions to contribute to poverty reduction, gender equity, and increased productivity, among other benefits (World Bank, 2021). Studies have shown that access to clean water and sanitation can improve productivity, reduce healthcare costs, and reduce the time burden on women and girls who are often responsible for water collection and decrease the risk of sexual violence (World Bank, 2021; Clasen et al., 2014).

Overall, the distribution of journals in this dataset suggests that evaluations of WASH interventions in rural SSA are being approached from a multidisciplinary perspective mainly when addressing health. However, they still require recognition of the environmental, and development-related challenges in the region and considerations about the practicalities of conducting the evaluations.

C. PRODUCTIVITY AND PERFORMANCE OF THE JOURNALS

One insight that can be gleaned from the productivity and performance data of the journals is the limited representation of developing countries in the literature on WASH interventions in rural SSA. While Ethiopia, Ghana, and Nigeria were among the top countries focused upon in the dataset, they account for only 3% of the journal country origin each. This suggests that there may be a gap in research capacity and funding in these countries, as well as potential barriers to publication and dissemination of research findings.

Furthermore, the fact that the only two journals originating from developing countries have no journal rankings or information about their global standing is significant. This suggests that there may be a lack of infrastructure or resources available to these journals to compete on a global scale. It also highlights the potential for bias towards research conducted in developed countries, as they may have greater access to funding and resources for research.

Regarding the distribution of publications across journals, the concentration of articles in just three journals (American Journal of Tropical Medicine and Hygiene, PLOS ONE, and International Journal of Environmental Research and Public Health) was noteworthy. These journals were all highly regarded and have achieved JCI scores indicating that their published papers are cited at or above the world citation average in their subject categories.

However, it is also important to note that only 47% of publications were published in journals with a JCI equal to or above 1, indicating that a lack of recognition for some of the research may be due to the limited availability of high-quality journals that specialize in WASH interventions in rural SSA. The majority of the publications may be published in general public health or environmental science journals, which may not prioritize research on WASH interventions or LMIC settings such as those in rural SSA.

Additionally, only 16% of publications were published in journals with a JCI above 2, further underscoring the potential misrecognition of or underinvestment in research on WASH interventions in LMIC's.

6.1.3 PUBLISHING AUTHORS

A. COUNTRY OF ORIGIN OF AUTHOR AFFILIATIONS

The distribution of author organizations highlights the dominance of the USA in the field of development WASH interventions in rural SSA. The high representation of organizations from the USA could be due to various reasons, such as funding opportunities, institutional support, and established research networks. Another possible insight would be that this dominance may reflect the country's strong influence and funding in the field of WASH interventions in SSA, as well as its academic and research capacities. This may also suggest a potential imbalance in the global distribution of resources and expertise in this area, with limited participation and representation from developing countries.

The low representation of organizations from SSA and other developing regions mirrors the findings in the above. Furthermore, the concentration of author organizations from SSA in a few countries such as Kenya, Zimbabwe, and Ethiopia may indicate the need for more equitable distribution of resources and support for research in other countries in the region.

Additionally, this may indicate the existence of knowledge hubs or research clusters in these locations, which may benefit from further collaboration and knowledge sharing with other regions.

B. PRODUCTIVITY AND PERFORMANCE OF THE AUTHORS

The first insight that can be drawn is that there was a strong correlation between the productivity of authors and the number of studies written on the same trials. This is particularly evident from the fact that 17 out of the top 20 most productive authors in the dataset are from the research teams of the SHINE trial and the WASH Benefits Kenya trial, which have the highest number of studies written on them. This suggests that there may be a trend towards more collaborative research within specific trials or research programs, which can lead to higher productivity for authors involved in those programs.

On the same note, the second insight is related to the field of interest of the most productive authors. The fact that all four of the most productive authors are concerned with subjects related to tropical medicine and health may reinforce the predominance of the notion of the intersectionality of the development field of water supply and that of health in today's research.

Furthermore, it is worth noting that the two most productive authors, Humphrey JH and Ntozini R, were both affiliated with the Zvitambo Institute for Maternal and Child Health Research in Zimbabwe. This suggests that the

SHINE trial, which is a collaborative project between the United States and Zimbabwe, has been highly productive in terms of producing research output. Similarly, Pickering AJ, who is a recurrent author in studies related to the WASH Benefits Kenya trial, is affiliated with the University of California, which is a partner institution in this collaborative project between the United States and Kenya. This may highlight the importance of international collaborations in conducting research in developing countries and the potential benefits of such collaborations in terms of producing research output.

Moreover, the fact that the top 20 most productive authors are largely affiliated with the research teams of the two trials with the highest number of studies may indicate that a few research projects can have a significant impact on the research output in a given field. This may underscore the importance of large-scale, well-funded, and collaborative research efforts in driving progress and producing knowledge that can inform policy and practice.

Overall, the findings suggest that international collaborations and investments in research in relevant fields of public health in developing countries can contribute to the production of high-quality research output, which can be important in advancing knowledge and informing policies and interventions to improve health outcomes in these contexts. However, and with a dataset of 30 papers, it is important to note that the findings and insights drawn from the analysis were limited in their scope and may not necessarily be generalizable to the broader field of global health or even to other subfields within global health.

6.2 NETWORK ANALYSIS

Firstly, the fact that 25 out of the 30 publications in the dataset have cited each other at least once suggests that there is a high level of interconnectivity and interdependence among these publications. This was a first indication that the authors in the dataset were working on closely related topics or on long-term research focus that has evolved over time.

This may also suggest that the authors have developed a significant body of knowledge on the topic and that their research has been influential within their field, or at the very least the miniature field of the dataset. However, it is important to note that citing one another does not necessarily indicate the quality or significance of the research, but rather reflects the authors' recognition of the work done by their colleagues.

Upon closer inspection of the reason why the papers in the dataset were cited by each other, it was found that they mostly do so to validate the results they obtained. This may suggest that the authors were building on each other's work and using it to support their own research, which can be a positive indication of a collaborative and productive research community.

However, it is important to note that this high level of interconnectivity and interdependence within the dataset could have also led to potential biases and limitations in the findings, as the authors may have a shared perspective or approaches that could skew the results in a particular direction and narrow the scope of the research.

Another thing to note is that the high level of interconnectivity and interdependence among the publications may have affected the representativeness of the H-index as a metric for evaluating the productivity and impact of the authors. The H-index is based on the number of publications and the number of citations that these publications receive, but if a large proportion of the citations are from other publications within the dataset, which

could represent a niche research community, this may artificially inflate the H-index. Therefore, it was important to consider other factors when evaluating the productivity and impact of the authors, such as the quality and significance of their research and the broader impact that it has had within their field.

Furthermore, it was found that the most cited papers by external literature in the network, mainly (Null et al., 2018; Humphrey et al., 2019), are related to a topic that is also the most common outcome studied in the dataset, which is diarrhea prevalence, it may suggest that the dataset is representative of the broader research landscape within the general research field.

This alignment between the external literature and the dataset may indicate that the research being conducted within the dataset is relevant and considerably recognized by the broader external research community, and that it is addressing an important health concern that is of interest to a wide range of researchers.

In addition, the fact that four of the studies in a dataset of 30 publications were found to have focused on the same brand of filter, the LifeStraw hollow-fiber filter, raised questions about the diversity of the products being evaluated in the field, or some potential bias towards this particular brand or type of filter.

On the same note, three out of the five publications that were cited by no other study in the dataset, while having no perceivable disparities with the other publications. This may suggest that there is a potential tendency within the research community of the dataset to favor certain approaches or interventions, potentially leading to a lack of diversity in the research being conducted or incurred bias in the findings.

Lastly, three major themes were identified in the dataset, one pertaining to the SHINE and WASH Benefits Kenya trials, one to safe water storage and prevention of recontamination and their effects on diarrhea, and one on POU household water treatments and commitment, grossly match the clustering obtained in the citation analysis. The different themes identified are rather limited in number, and mostly similar in the overarching theme, especially the last two, which could be another indication of the thematic homogeneity of the dataset and its close interconnectedness.

6.3 FOCUS OF THE DATASET

The extensive use of RCT's as an evaluation method was a notable characteristic of the publications considered in this study. The search for impact evaluations of interventions in the water sector of rural SSA yielded a dataset mostly comprised of RCT's related to household-level WASH interventions. Focusing on small or mid-level scale interventions mirrors an emergent trend identified by literature as prioritizing poverty alleviation over (and in support of) development, and as being indicative of the prevalent movement in development research and policy that are shifting away from viewing development as structural transformation. Indeed, the mainstream larger emphasis of "thinking small" is seen to marginalize issues concerning large centralized projects, and to focus on identifying the most effective methods for implementing small-scale, mostly technical, interventions (Kvangraven, 2019). This shift in global approaches has given rise to the popularity of RCT's, which have been reported to be increasingly used in the past 10 years. The main points featured in the debate over their use, which were elaborated upon in Chapter 4, can be clearly identified in within the dataset of this study.

STANDARDIZATION

In the dataset of this study, the included trials exhibited a certain level of standardized design features. These included the trial settings which were mostly household-based, the trial designs in terms of blinding and randomization, as well as the population they targeted. Likewise, the type of targeted outcomes, which had identical definitions set by global international organizations such as the World Health Organization (WHO) was recurrent. The data collection methods utilized were also recurring.

That however, was not applicable to other equally important aspects of the trial designs such as the durations of the trials, the type of baseline conditions data, and the frequency of data collection and household visits thereby leading to different recall times of self-reported outcomes. Similarly, comparing the results of trials with different experiment strategies, essentially those that did not collect data on how well the intervention was functioning or those that relied solely on self-reported data without any observational parameters, may also be questionable.

GENERALIZABILITY OF THE FINDINGS

The generalizability of the findings of certain trials in the dataset may have also been affected by the potential bias incurred by their designs and/or overall approaches. The design features include the duration of the trials, the pre-determination of the baseline conditions, the insufficiency of disaggregation of the data by gender and/or socio-economic indicators during the baseline data collection, the restriction of certain inclusion and exclusion criteria of the participants, the lack of blinding in almost all of the trials, the reliance on self-reported or indirectly-inferred outcomes, and the absence of efforts to investigate the reasons or underlying mechanisms that lead to those findings. As for the approaches adopted by the trials, it was noted that they may have failed at simulating sufficiently real-world conditions, by opting for the free hand-in of interventions for example, the swift repair or replacement of defected hardware, and the frequent health messaging and household data collection visits.

Furthermore, the recurrent patterns in the elements of the trials identified, including the designs and methodologies of the trials, and the targeted outcomes, along with the findings deducted from the bibliometric and network analyses related to the dataset research community being tightknit, may also reveal a certain degree of compromise in the generalizability of the findings of the dataset studies.

One thing to note however, is that the use of the 3ie portal to identify relevant studies may have preset constraints on the generalizability potential of the studies. Indeed, although the 3ie is a widely used source for evidence on international development interventions, it is important to note that the organization was founded by a group of donor countries and international institutions, and its funders are primarily from the Global North. This may introduce potential partiality towards certain types of interventions or research questions, as well as limit the scope of the studies available on the portal. Additionally, it is important to consider the potential language and publication bias on the portal, as it primarily includes publications in English and may have excluded relevant studies in other languages or published in non-peer-reviewed sources. On the same note, the dataset included a limited number of publications from the Global South, which may have resulted in a biased perspective that does not reflect the realities and contexts of the region and the way the interventions were perceived and appraised by the local communities. Moreover, the fact that the publications were only in English may have resulted in the exclusion of potentially valuable and more diverse studies conducted in other languages, leading to an incomplete or limited analysis and generalizability of the findings.

Overall, while the findings of the dataset RCT's may be valuable, it is important to consider the limitations and critical insights discussed above to ensure the generalizability and transferability of the interventions to other populations and contexts. Furthermore, by considering all of the aforementioned points, the debate centering around the idealized perception in literature of RCT's being the "gold standard" of clinical trial research presents valid arguments in terms of RCT's being prone to the potential compromise of their promise of optimal reduction in bias and the generalizability of their findings. This is at least was demonstrable and applicable to the dataset considered in this study.

DATASET REPRESENTATIVENESS

Finally, in terms of dataset representativeness, this study deemed its content and findings to be consistently comparable to the general external literature. This observation was concluded in part from the bibliometric representativeness analysis performed, which involved the comparison of the bibliometric information of the most recurrent constituents of the publications in the dataset with those identified for a dataset extracted from WoS, which was meant to represent external literature.

Likewise, it was found that the studies within and outside of the dataset share several commonalities, including the use of the preference of RCT design, difficulties with blinding, and a focus on low-cost water treatment hardware. Moreover, interventions targeting water treatment were more prevalent than other interventions in both studies. Finally, discrepancies in study results were also observed in studies conducted outside of the dataset, highlighting the need for caution in generalizing study findings.

Conclusion

In response to the growing recognition of the importance of evidence-based research in international development research, on-the-ground interventions as well as international decision and policymaking, this study aimed to explore the field of foreign aid evaluations of interventions in the water sector of rural Sub-Saharan Africa (SSA). To do so, the study centered around a set of publications extracted from the International Initiative for Impact Evaluation (3ie) Development Evidence Portal (DEP), one of the largest repositories in the world for evidence-based research in low and middle income countries. Data from the different studies in the dataset pertaining to the characteristics of the evaluations and trials were first collected. Bibliographic and network analyses were then conducted in order to identify the structure of the research community formed within the dataset. The content of the dataset papers was then analyzed and situated within the external literature in order to validate and strengthen the conclusions drawn from the analysis, contextualize the findings and recommendations from the study, and identify gaps and areas for further research.

FINDNGS OF THE STUDY

OVERVIEW OF THE DATASET CHARACTERISTICS

In broad terms, the dataset comprised 30 publications all found to be impact evaluations of randomized controlled trials (RCT's) related to Water, Sanitation, and Hygiene (WASH) interventions, spread around nine countries with the majority located in Kenya. The duration of the trials varied widely, with some as short as 8 weeks and the longest being 29 months.

As for the focus of the trials, the dataset featured only four trials that explored the effects of improved sanitation, three of which were large-scale three large-scale trials: the Sanitation, Hygiene, Infant Nutrition Efficacy (SHINE) trial in Zimbabwe, the WASH Benefits Kenya trial in Kenya and the '*Villages et Ecoles Assainis*') (VEA) in the Democratic Republic of Congo (DRC). The first two trials had several papers revolving around them in the dataset, with the first having seven publications and the second five. The trials investigated the effectiveness of nutritional supplementation coupled with improved water and sanitation facilities. As for the rest of the interventions, they mostly revolved around household water treatment hardware, with the most common being chlorination, followed by filtration, solar disinfection and lastly flocculent-disinfection. Two other types of interventions were included in two separate trials, which were improved water supply with borehole drilling and improved water storage and transportation.

Almost all of the trials investigated health-related outcomes for children under five years, with diarrhea prevalence being by far the most recurrent studied outcome. In case of water treatment interventions, some of the trials also explored changes in water quality. Lastly, only one trial studied end-user preference of household water treatment hardware.

In terms of results, significant heterogeneity among the trial findings was observed, especially those pertaining to diarrhea prevalence, which may suggest that the level of intervention effectiveness may depend on a variety of conditions that research to date cannot or have not fully explained yet.

STRUCTURE OF THE RESEARCH COMMUNITY

As for the structure community, the study revolved around a dataset that exhibited some key disparities in the distribution of trials and studies across SSA with the majority aggregated in Kenya, and the bulk of studies being conducted by organizations and institutions from the Global North and published in journals mainly oriented towards the Global North audience. Concurrently, within the Global North-dominated research community in the dataset, mostly publications were found to cite each other and the same sources frequently, indicating a research community whose constituents are well interconnected and that build on each other's contribution.

Furthermore, after the categorization of the content and characteristics of the dataset publications, the studies were found to employ similar approaches to investigating outcomes that were highly recurrent, mainly the prevalence of children diarrheal episodes. All of these observations indicated a potential lack of diversity in the knowledge pool. Accordingly, and as researchers draw from the same body of literature, a narrow focus on a particular set of interventions or approaches may have prevailed, thereby limiting the potential for innovation and experimentation. Additionally, the lack of representation from local researchers and organizations in the dataset and the overrepresentation of organizations from the Global North could perpetuate a power dynamic where research priorities and methods are set by those outside the community being studied, further limiting diversity in approaches and knowledge. This may also have contributed to the lack of significant intervention effects on the uptake and improvement in the investigated outcomes.

DEDUCTED FINDINGS AND LITERATURE GAPS

Firstly, the study found that most of the publications in the dataset investigated WASH interventions, with the majority related to the provision of low-cost household water treatments. This may highlight the growing recognition by the development research community of the importance of addressing water and sanitation challenges in low-income countries as a means of improving health outcomes and reducing poverty.

Furthermore, although implicitly conveyed, this focus on low-cost interventions may reflect a recognition of the financial constraints faced by many individuals and families in low-income countries, and the need for interventions that can be implemented on a wide scale. It may also reveal a shift towards a more decentralized approach to water provision. Instead of targeting point of source (POS) water, which can be costly and difficult to maintain due to the logistical, financial, institutional and political constraints that the communities may face, household water treatments have the ability to provide a simple and affordable solutions that may afford households the ability to govern the state of their water independently thereby enhancing their water security.

Secondly, all the publications in the dataset were impact evaluations of RCT's revealing the current research community's preference for this particular experiment design, which was usually reported by the studies in the dataset to be used because of its ability to establish causal relationships between interventions and outcomes. Indeed, both in the external literature and in the papers in this dataset, RCT's were often reported to be utilized because of their aptitude to minimize bias by permitting to isolate the effect of the intervention from other factors

that may be affecting the outcomes through the randomization of the participants between control and intervention groups. This made it possible to have an indication of whether the intervention was responsible for any observed changes in the outcomes

However, possibly defeating the reason to be chosen, RCT's in the dataset were found to include factors that compromise their promise of minimizing bias. Almost all RCT's were non-blinded trials with several reporting to fail in the randomization process, which reveals potential limitations in the quality and rigor of the studies in the dataset. The limitations may be due to the substantial financial disbursement and time required to conduct rigorous RCT's, which may be exacerbated by the limitedness of the studied rural areas in terms of resource and infrastructural bases.

Furthermore, in the context of the objectives of the studies in the dataset, the recurrent approach adopted by the trials may be questionable due to their potential inability to simulate the functioning of an intervention in "real-world" settings. That is because the trials featured in the dataset opted for the free provision of hardware with (or without) recurrent messaging. This may have been exacerbated by the trials not reporting on investigating the cost-effectiveness of the interventions, therefore potentially misrepresenting the results they obtain for recipient-uptake results. This may be further amplified in the cases where the trial duration was insufficient and may not have accounted for prevalent seasonal variations. This is especially important as the effectiveness of the intervention and the user uptake rate have been shown by the studies in the dataset to change over time. This may be indicative of the need to explore alternative or complementary research designs that address the limitations of RCTs and provide a more comprehensive evaluation of interventions in real-world settings.

Thirdly, this study revealed that the research community, despite being tightknit, faces persistent challenges in conducting rigorous studies that minimize bias. The fact that the same set of study limitations were recurrent across the dataset suggested that there are systematic issues in the way these studies are being designed and conducted. For example, the concern about disease contraction from other contamination pathways highlights the complexity of addressing WASH challenges in rural SSA settings. Similarly, issues in terms of the potential bias incurred from self-reported data and unblinded trials were also persistent. This underscores the need for more comprehensive, intensive and/or higher quality interventions that undertake more innovative approaches.

REPRESENTATIVENESS OF THE DATASET

The dataset was found to consistently mirror the general literature in the development field. Firstly, the different bibliometric constituents and the structure of the research community formed by the studies in the dataset were consistent with those in the external literature. These findings were concluded through the bibliometric representative analysis done that compared the most recurrent bibliometric constituents of the dataset papers with those of a dataset formed from WoS meant to represent the content of the external research field.

Secondly, the studies within and outside of the dataset were found to share several commonalities, including the use of the RCT design, difficulties with blinding, and a focus on low-cost water treatment hardware. Moreover, interventions targeting water treatment were more prevalent than other interventions in both studies. Finally, discrepancies in study results were also observed in studies conducted outside of the dataset, highlighting the need for caution in generalizing study findings.

In general, although the outcomes of this study have restricted applicability, they offer some perspectives into the scholarly work and influence of a small representative fraction of the field of developmental aid in the water sector and its assessment. The outcomes and findings drawn underscore the importance of sizable and cooperative research projects, as well as the requirement for sustained support towards improving research and implementation in low- and middle-income nations.

The aim was to identify gaps in knowledge and areas for future research, as well as to provide recommendations for improving the quality and impact of future evaluations in this field. By conducting a thorough analysis of the research community and the evidence-based research available, this study contributes to the ongoing effort to improve the effectiveness of foreign aid interventions in rural Sub-Saharan Africa, and to ensure that interventions are based on sound evidence and lead to positive outcomes for the communities they aim to serve.

STUDY LIMITATIONS

One limitation of this study is the use of a narrow set of exclusion criteria, which may have resulted in the exclusion of that could have contributed to a more comprehensive overview and analysis. For examples, studies that assessed interventions related solely to education, behavioral change, awareness campaigns, or the exclusive provision of soap were eliminated, which may have limited the scope of the analysis and excluded potentially valuable studies that could have provided insights into alternative approaches to water interventions and their evaluations in the region.

Moreover, the small sample size of the dataset is also a limitation of the study. Only 30 papers were included in the analysis, which may not be representative of the broader literature on the subject. The skewed distribution of the publications, with some trials having more publications than others, may have also influenced the findings of the study and introduced bias.

Lastly, the study did not examine the quality of the publications included in the dataset, such as the rigor of the study design or the validity of the data analysis methods. Therefore, the conclusions drawn from the study may be limited by the quality of the publications analyzed.

RECOMMENDATIONS

This study has drawn several recommendations that may help in addressing some of the identified observations and gaps in literature:

1. Data collection would be better extended to understudied quality of life and socio-economic outcomes such as health, empowerment, safety, household income, and long-term wage earnings resulting from WASH improvements experienced during childhood;
2. Longer-term measurement of adherence and quality of life improvements are necessary;
3. The feasibility, acceptability, cost-effectiveness, and sustainability ought to be more adequately assessed
4. More evaluations collecting data on objective measures of gendered behavior change, health, and socioeconomic outcomes are needed;

5. The use of natural experimental approaches should may be considered and assessed in their ability to control for unobservable confounding without distorting the natural process of intervention rollout or using outcome data collection methods that may introduce reporting biases;
6. Impact evaluations of interventions targeting or presenting disaggregated data for vulnerable populations, particularly over the life-course and for people living with a disability, should be given attention;
7. The publication of research from developing countries in international journals should be encouraged and supported to promote the representation and visibility of diverse voices in the global research community.
8. Prioritizing economic considerations in the studies of the evaluation of WASH interventions, including the costs of conducting evaluations, is essential to allocate limited resources efficiently and promote the sustainability and scalability of WASH interventions for greater impact on health and economic development outcomes.

APPENDIX A

Table 8: Details of the trials with multiple studies in the dataset

Trial Name	Studies of the trial	Description of the trial
Sanitation Hygiene Infant Nutrition Efficacy (SHINE)	(Humphrey et al., 2015); (Chasekwa et al., 2018); (Humphrey et al., 2019); (Gladstone et al., 2019); (Chandna et al., 2020); (Gough et al., 2020)	The SHINE trial aimed to evaluate the effects of two interventions on infant health and stunting in rural Zimbabwe. The trial was community-based and cluster-randomized, with two interventions being tested independently and combined: protecting infants from fecal ingestion through a WASH intervention and optimizing their nutrition through an Infant and Young Child Feeding (IYCF) intervention. The trial measured two causal pathways and clusters were randomly assigned to standard of care, IYCF, WASH, or both IYCF and WASH. The IYCF intervention included the provision of a lipid-based nutrient supplement and complementary feeding counseling, while the WASH intervention involves the construction of improved latrines, provision of handwashing stations and hygiene counseling.
WASH Benefits Kenya	(Christensen et al; 2015); (Null et al., 2018); (Stewert et al; 2018); (Swarthout et al, 2020)	In the WASH Benefits Kenya trial, clusters that were geographically-adjacent were randomly allocated to three groups: active control, passive control, or compound-level interventions. The compound-level interventions involved the building of new (or rehabilitating existing) latrines, household visits to promote target behaviors, and the restocking treatment hardware. The households were then assigned to one of eight study arms, including interventions such as drinking chlorinated water, safe sanitation, handwashing with soap, and a combination of water, sanitation, and handwashing. Additionally, counseling was provided on appropriate maternal, IYCF, and small-quantity lipid-based nutrient supplements were given from 6-24 months. Some clusters received the combined intervention of water, sanitation, handwashing, and nutrition. Data collection was also conducted in passive control clusters, while active control clusters only received household visits to measure stunting indicators, mid-upper-arm circumference.

CARE Kenya	(Alber et al., 2010); (Luoto et al., 2014)	The CARE trial was conducted into two settings and countries: in rural Kenya and the “urban slums” in Bangladesh. Both locations received almost identical interventions, which went as follows: after the introduction of three household point of use water treatment products, namely chlorination (WaterGuard bottles), flocculation disinfectants, and ceramic filters, enumerators randomly assigned one of the products to respondents for a two-month trial. Half of the respondents were presented with a "positively framed" message that advocated the benefits of contamination-free water, while the other half received a "contrast framed" message that cautioned about the health risks of contaminated water. Furthermore, enumerators asked half of the respondents to commit verbally to using their assigned product for all of their drinking water and provided a poster as a reminder. The assignment of products and marketing treatments were implemented orthogonally. All marketing treatments emphasized the expected health benefits of using any safe water product and did not present relative comparisons of one product versus another. After two months, enumerators revisited all households to collect stored water and test for product usage, then delivered the same marketing treatment or treatments. Enumerators then distributed one of the remaining POU products for a new two-month trial and collected any leftover supplies of the previously assigned product. This process was repeated until every participant in the treatment group had experienced a two-month trial of each of the three POU products in random order. Only one difference exists in the intervention designs of the two countries, mainly that intervention households in Bangladesh did not receive safe storage containers during the chlorination and flocculent-disinfectant periods as they preferred to use traditional storage devices. For this trial, (Albert et al., 2010) assessed end-user preference of water treatment product while (Luoto et al., 2014) investigated changes in intervention uptake by the participants with respect to the different types of “nudges” implemented.

APPENDIX B

Table 9: Dataset Studies Characteristics and Classifications

Country	Title	Year	Duration	Interventions	Type of technology	Nature of intervention	Outcome(s)
DRC	(Quattrochi , 2021)	2021	5 months	Community mobilization, participatory planning, and capacity building + Subsidies + Health messaging	POS water supply: water points combined with sanitation and hygiene	Demand	Time and quantity of water transport| Water treatment and storage use| Hygiene and sanitation practices| Diarrhea morbidity| cost of water| school attendance
Kenya	(Fagerli, 2020)	2020	12 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity
Ethiopia	(Solomon, 2020)	2020	4 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity
Kenya	(Swarthout , 2020)	2020	20 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies| Combined hygiene with water quality and(or) sanitation	Both	ARI
Zimbabwe	SHINE: (McQuade, 2020)	2020	29 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies, children playground| Combined hygiene with water quality and(or) sanitation	Both	Diarrhea morbidity| Mortality
Zimbabwe	SHINE: (Gough, 2020)	2020	29 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene:	Both	EED| Stunting| Nutrition| Mortality

					handwashing supplies, children playground| Combined hygiene with water quality and(or) sanitation		
Zimbabwe	SHINE: (Gladstone, 2019)	2019	29 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies, children playground| Combined hygiene with water quality and(or) sanitation	Both	Stunting| Nutrition| Mortality
Zimbabwe	SHINE: (Humphrey, 2019)	2019	29 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies, children playground| Combined hygiene with water quality and(or) sanitation	Both	Stunting| Nutrition| Mortality
Kenya	Kenya (Stewart, 2018)	2018	29 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies| Combined hygiene with water quality and(or) sanitation	Both	Stunting| Nutrition
Zimbabwe	SHINE: (Chasekwa, 2018)	2018	20 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies, children playground| Combined hygiene with water quality and(or) sanitation	Both	Diarrhea morbidity| Stunting| Nutrition| Mortality
Kenya	Kenya (Null, 2018)	2018	29 months	Direct supply + Health messaging	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies| Combined hygiene with	Both	Diarrhea morbidity| Stunting| Nutrition| Mortality

					water quality and(or) sanitation		
Kenya	(Morris, 2018)	2018	20 months	Free hardware supply	POU water treatment: household supply	Supply	Diarrhea morbidity| ARI| Cryptosporidium| Mortality
Zimbabwe	SHINE: (Humphrey, 2015)	2015	4.5 months	Free hardware supply	POU water treatment: household supply| Sanitation: household latrines| Hygiene: handwashing supplies, children playground| Combined hygiene with water quality and(or) sanitation	Supply	Diarrhea morbidity| Stunting| Nutrition| Mortality
Ghana	(Cha, 2015)	2015	29 months	Direct supply + Health messaging	POS water supply: boreholes combined with hygiene	Both	Water treatment and storage use|Hygiene practices| Diarrhea morbity
Kenya	(Christensen, 2015)	2015	2 months	Direct supply + Health messaging	POU water treatment: household supply|Sanitation: household latrines| Hygiene: handwashing supplies for household use	Both	Water treatment and storage use| Construction, maintenance and use of latrines| Hygiene and sanitation practices
Kenya and Bangladesh	(Luoto, 2014)	2014	6 months	Direct supply + Health messaging	POU water treatment: household supply	Both	Water treatment and storage use
Rwanda	(Rosa, 2014)	2014	8 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use
Ethiopia	(Mengistie, 2013)	2013	5 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity| Mortality
Benin	(Gunther, 2013)	2013	4 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity
Ghana	(Stauber, 2012)	2012	4 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity

Kenya	(Patel, 2012)	2012	15 weeks	Direct supply + Health messaging	POS water supply: water points combined with sanitation and hygiene	Both	Water treatment and storage use| Diarrhea morbidity| ARI| Hygiene and sanitation practices
Kenya	(du Preez, 2011)	2011	12 months	Free hardware supply	POU water treatment: household supply	Supply	Diarrhea morbidity| Mortality
Kenya	(Kremer, 2011)	2011	17 months	Free hardware supply	POS water treatment: household supply	Supply	Water treatment and storage practices| Diarrhea morbidity| Nutrition | Stunting| Willingness to pay
DRC	(Boisson, 2010)	2010	15 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity| Mortality
Kenya	(Albert, 2010)	2010	6 months	End user preference	POU water treatment: household supply	Demand	Water treatment and storage use| End-user preference
Ethiopia	(Boisson, 2009)	2009	5 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity
South Africa and Zimbabwe	(du Preez, 2008)	2008	6 months	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity
Kenya	(Garrett, 2008)	2008	8 weeks	Community mobilization, participatory planning, and capacity building + Health messaging	POS water supply: water points combined with sanitation and hygiene	Demand	Water treatment and storage use| Diarrhea morbidity
Kenya	(Crump, 2005)	2005	20 weeks	Free hardware supply	POU water treatment: household supply	Supply	Water treatment and storage use| Diarrhea morbidity| Mortality

REFERENCES

Ain, Q.-ul, Riaz, H., & Afzal, M. T. (2019). Evaluation of H-index and its citation intensity based variants in the field of mathematics. *Scientometrics*, *119*(1), 187–211.

Bain R., Cronk R., Hossain R., Bonjour S., Onda K., Wright J., Bartram J. (2014). Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Medicine and International Health.* 19(8):917–927.

Baker, K., Butterworth, J., & Gichohi-Wainaina, W. N. (2019). Why is water still managed as a sector? Exploring prospects for integrated water resources management in Kenya. *Water Alternatives*, 12(2), 307-324.

Bastos Lima, M. G., Oria, P. A., & Dorea, C. C. (2015). Effectiveness of a community-based intervention to improve maternal and newborn health in the Northeast of Brazil. *Revista Brasileira De Ginecologia E Obstetrícia*, 37(8), 363-370.

Bauer, P., Kohne, K., & Kohne, K. (2016). Evaluating the generalizability and validity of study results. In Basic principles and practice of clinical research (pp. 281-291). *Springer*.

Bhutta, Z. A., & Victora, C. G. (2005). Scaling up interventions for improving health: the role of randomized trials. *The Journal of Clinical Investigation*, 115(10), 2842-2847.

Biran, A., Jenkins, M. W., & Dabrase, P. (2011). Public toilets and their customers in low-income Mumbai. *Environment and Urbanization*, 23(2), 415-432.

Biran, A., Schmidt, W. P., Wright, R., Jones, T., Seshadri, M., Isaac, P., ... & Curtis, V. (2012). The effect of a soap promotion and hygiene education campaign on handwashing behaviour in rural India: a cluster randomised trial. *Tropical Medicine and Internatonal Health*, 17(3), 312-321.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892-895.

Bornmann, L., & Leydesdorff, L. (2013). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Journal of the Association for Information Science and Technology*, 64(1), 95-109.

Bovens, M., Goodin, R. E., & Schillemans, T. (2016). The Oxford Handbook of Public Accountability. *Oxford University Press*.

Caruso, B. A., Clasen, T. F., Hadley, C., Yount, K. M., Haardörfer, R., Rout, M., ... & Dreibelbis, R. (2020). Understanding and defining sanitation insecurity: women's gendered experiences of urination, defecation and menstruation in rural Odisha, India. *BMJ Global Health*, 5(9), e003074.

Chandrasekhar, S., Ostermann, J., Wiegandt, A., Selleng, E., Sarfo, F., Boahen, O., et al. (2016). Uptake and sustainability of a water, sanitation and hygiene intervention in rural Ghana. *Journal of Water, Sanitation, and Hygiene for Development*, 1;6(4):721.

Chow, S. C., Shao, J., & Wang, H. (2018). Sample Size Calculations in Clinical Research. Boca Raton, FL: *CRC Press*.

Cissé, G., Kié, R. G., Koné, B., & Koné, D. (2016). Water supply in sub-Saharan Africa: The history and challenges of a centralised sector. *Water Alternatives*, 9(1), 28-48.

Cope, H., Lucas, P., Whiting, K., & Schmidt, W. P. (2018). Ethical challenges of conducting randomized controlled trials in rural communities: the case of a WASH intervention study in Tanzania. *Journal of Empirical Research on Human Research Ethics*, 13(3), 285-294.

Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, 1(3), 193-203.

Crocker, J., Saywell, D., & Bartram, J. (2017). Sustainability of community-led total sanitation outcomes: Evidence from Ethiopia and Ghana. *International Journal of Hygiene and Environmental Health*, 220(4), 551-557. doi: 10.1016/j.ijheh.2017.01.006

Devereaux, P. J., Manns, B. J., Ghali, W. A., & Quan, H. (2001). The need for randomized control trials in surgical outcomes research. Journal of Surgical Research, 109(2), 244-255. Kjaergard, L. L., Nikolova, D., & Gluud, C. (2001). Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology*, 33(6), 1194-1199.

Djulbegovic, B., & Guyatt, G. H. (2017). Progress in evidence-based medicine: a quarter century on. *The Lancet*, 390(10092), 415-423.

Dreibelbis, R., Freeman, M. C., Greene, L. E., Saboori, S., & Rheingans, R. D. (2013). The impact of school water, sanitation, and hygiene interventions on the health of younger siblings of pupils: a cluster-randomized trial in Kenya. *The American journal of tropical medicine and hygiene*, 89(5), 821-828.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., . . . Wechsler, H. (2018). Youth risk behavior surveillance - United States, 2017. *Morbidity and Mortality Weekly Report*, 67(8), 1-114.

Elsevier. (n.d.). What is a Citation Count? Scopus. https://www.elsevier.com/solutions/scopus/how-scopus-works/content/citation-count

Emanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *Jama*, 283(20), 2701-2711.

Fisher, J., Cook, A., & James, S. (2018). Implementing water quality and hygiene interventions: challenges and opportunities. *Journal of Water, Sanitation and Hygiene for Development*, 8(2), 224-235.

Foster, T., Willetts, J., Lane, M., Thomson, P., Katuva, J., Hope, R., ... & Mbalo, S. (2018). Risk factors associated with rural water supply failure: A 30-year retrospective study of handpumps on the south coast of Kenya. Science of the Total Environment, 626, 156-164.

Freedman, B. (1987). Equipoise and the ethics of clinical research. *New England Journal of Medicine*, 317(3), 141-145.

Freeman, M. C., Stocks, M. E., Cumming, O., Jeandron, A., Higgins, J. P., Wolf, J., Prüss-Ustün, A., & Bonjour, S. (2017). Systematic review: hygiene and health promotion interventions to prevent diarrhoea in lower-income countries. *Tropical Medicine & International Health*, 22(11), 1370-1387.

Gagolewski, M., & Mesiar, R. (2012). Aggregating different paper quality measures with a generalized h-index. *Journal of Informetrics*, *6*(4), 566–579.

Gautam, O. P., Schmidt, W. P., Cairncross, S., & Cavill, S. (2010). Trial of a novel intervention to improve multiple food hygiene behaviors in Nepal. *American Journal of Tropical Medicine and Hygiene*, 83(3), 657-666.

George, C.M., Oldja L, Biswas, S., et al. (2015). Geophagy is associated with environmental enteropathy and stunting in children in rural Bangladesh. *American Journal of Tropical Medicine and Hygiene*, 92(6):1117-1124.

Gill, C. J., Hyder, A. A., & Prabhakaran, D. (2013). Strengthening the ethics of global health research. *Bulletin of the World Health Organization*, 91(4), 243-243A.

Glänzel., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management*, 35(1), 31-44. doi: 10.1016/S0306-4573(98)00058-8.

Glänzel, W. (2006). On the Opportunities and Limitations of the H-index. *Science Focus* (1).

Glasziou, P., Vandenbroucke, J., & Chalmers, I. (2004). Assessing the quality of research. *BMJ: British Medical Journal*, 328(7430), 39-41.

Gleick, P. H. (2003). Global freshwater resources: soft-path solutions for the 21st century. *Science*, 302(5650), 1524-1528.

Gnilo, M., Nitièma, J.A., Pilger, D., Coulibaly, S, Tarnagda, G., Soura, A.B., et al. (2020). The impact of a WASH intervention on behaviour change and health outcomes in rural Burkina Faso: a quasi-experimental study. *BMC Public Health*, 20(1):1-1.

Haider, B. A., Bhutta, Z. A., & Lassi, Z. S. (2019). Combined impact of multiple micronutrient supplementation on survival, growth, and morbidity of children under-five: a systematic review and meta-analysis. *BMC public health*, 19(3), 1-19.

Hanchate, S., Sundaram, N., Sivaprakasam, P., Jacks, G., & Ramani, K. V. (2019). Health equity in rural India: Evidence from a randomized controlled trial of sanitation and hygiene interventions. *International Journal for Equity in Health*, 18(1), 1-12.

Harcourt, W. (2012). Occupy Education: Living and Learning Sustainability. *New York: Autonomedia*.

Harzing, A. W. (2017). Citation analysis: Weeding out the bad apples. In R. P. Vásquez-Martínez (Ed.), Handbook of Research on Scholarly Publishing and Research Methods (pp. 238-256). *IGI Global*.

Hayes, R. J., Moulton, L. H., & Cluster Randomised Trials Working Group. (2016). Cluster Randomised Trials. *Boca Raton, FL: CRC Press*.

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *BMJ*, 350, h391

Higgins, J. P., & Green, S. (Eds.). (2011). Cochrane handbook for systematic reviews of interventions (Vol. 4). *John Wiley & Sons*.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., . . . Sterne, J. A. C. (2017). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 355, i5923.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572.

Horwitz, R. I., & Horwitz, S. M. (2017). What is the evidence for evidence-based medicine?. *BMJ: British Medical Journal*, 358, j3991.

Howard, G., & Bartram, J. (2003). Domestic Water Quantity, Service Level and Health. Geneva: *World Health Organization*.

International Initiative for Impact Evaluation (3ie). (n.d.). WASH Evidence Portal. Retrieved from https://www.3ieimpact.org/evidence-hub/programmes-and-approaches/wash

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124.

Kaptchuk, T. J., Miller, F. G., & Benedetti, F. (2018). Sham device v inert pill: randomised controlled trials and active comparators. *BMJ (Clinical Research Ed.),* 363, k4347.

Kazemi, F., Namdar, M., Ziapour, B., & Keshavarz Mohammadian, A. (2019). The role of mobile health in improving water, sanitation, and hygiene services: a systematic review. *Journal of environmental health science & engineering*, 17(1), 443-453.

Kenny, A., Laidlaw, J., & Wong, R. (2019). Evidence-based policy in development: What is it, and how do we get more of it? ODI Working Paper 557. *Overseas Development Institute*.

Kosek, M. N., Mduma, E., & Kosek, P. S. (2017). Prevention of stunting in children under 2 years old: common themes across four successful interventions. *The American journal of clinical nutrition*, 106(Supplement_2), 359S-364S.

Kvangraven, I. H. (2019). *Impoverished Economics? unpacking the economics Nobel prize*. *Open Democracy*. Retrieved March 12, 2023, from https://www.opendemocracy.net/en/oureconomy/impoverished-economics-unpacking-economics-nobel-prize/

Langlois, E. V., & Miszkurka, M. (2019). Before-and-after concurrent control studies. In Encyclopedia of Biostatistics (2nd ed.). *John Wiley & Sons, Ltd*.

Lathabai, H. H., Prabhakaran, T., & Changat, M. (2017). Contextual productivity assessment of authors and journals: a network scientometric approach. *Scientometrics*, *110*(2), 711-737.

Liu, X., Zhang, Y., & Yang, J. (2017). Does H-index correlate with academic rank of full-time urology faculty members? An objective evaluation of urology faculty research productivity using the H-index. *Urology*, 106, 76-81.

Mdege, N. D., Chindove, S., Ali, S., & Brown, J. (2019). Evidence synthesis for scaling up: Opportunities and challenges. *BMJ Global Health*, 4(8), e001785.

Miller, F. G., & Joffe, S. (2016). Equipoise and the duty of care in clinical research. *New England Journal of Medicine*, 374(14), 1365-1371.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., & Altman, D. G. (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *Journal of clinical epidemiology*, 63(8), e1-e37.

Molina-Montes, E., Wark, P. A., Sánchez, M.-J., Norat, T., Jakszyn, P., Luján-Barroso, L., ... & Bueno-de-Mesquita, H. B. (2018). Dietary intake of iron, heme-iron and magnesium and pancreatic cancer risk in the European Prospective Investigation into Cancer and Nutrition cohort. *International Journal of Cancer*, 142(6), 1154-1163.

Mosler, H.J. (2012). A systematic approach to behavior change interventions for the water and sanitation sector in developing countries: a conceptual model, a review, and a guideline. *International Journal of Environmental Health Research*, 1;22(4):431-49.

Nair, H., Simões, E. A., Rudan, I., Gessner, B. D., Azziz-Baumgartner, E., Zhang, J. S., ... & Weber, M. W. (2020). Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis. *The Lancet*, 381(9875), 1380-1390.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review,* 45(2), 167-256. doi: 10.1137/S003614450342480

OECD water - water. OECD. (n.d.). Retrieved September 18, 2022, from https://www.oecd.org/water/

O'Reilly, K., & Louis, E. (2014). The toilet tripod: understanding successful sanitation in rural India. *Health & Place*, 29, 43-51.

Overview. World Bank. (n.d.). Retrieved September 18, 2022, from https://www.worldbank.org/en/topic/water/overview

Papaioannou, T. G. (2018). Bibliometric indicators and analysis of research systems: Methods and examples. *Evaluation and Program Planning*, 69, 31-46.

Parker, A. H., Nagel, C. L., & Thomas, E. A. (2020). Time is of the essence: Understanding the burden of conducting randomized controlled trials in water, sanitation, and hygiene in low-resource settings. *International Journal of Environmental Research and Public Health*, 17(12), 4447.

Parpia, A. S., Ndeffo-Mbah, M., Wenzel, N. S., Galvani, A. P., & Meyers, L. A. (2016). Evaluating vaccination strategies for Zika virus in the Americas. *Annuals of Internal Medicine*, 165(9), 585-593.

Peletz, R., Simonsen, L., Wagner, E., Baisley, K., Jia, T., Keenan, J. D., ... & Pickering, A. J. (2016). Costs and cost-effectiveness of 9 interventions for improving the cleanliness of shared toilets in informal settlements in KwaZulu-Natal, South Africa. *PloS One*, 11(9), e0161737.

Peprah, D., Baker, K. K., Moe, C. L., Robb, K. A., Wellington, N., Yakubu, H., & Clasen, T. F. (2021). Use of telephone surveys for rapid monitoring of rural water, sanitation, and hygiene (WASH) coverage: a cross-sectional study in Ghana. *BMC public health*, 21(1), 1-10.

Peprah, D., Mohammed, A. H., Mensah, E. K., Karikari, T. K., & Abdul-Rahman, A. (2020). Factors influencing the implementation of clinical trials in Ghana: A qualitative study. *PloS one*, 16(3), e0248685.

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. *Blackwell*.

Pluskiewicz, W., Drozdzowska, B., Adamczyk, P., & Noga, K. (2019). Scientific quality index: A composite size-independent metric compared with H-index for 480 medical researchers.

Prüss-Üstün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M. C., ... & Hunter, P. R. (2019). Burden of disease from inadequate water, sanitation and hygiene in low-and middle-income settings: a retrospective analysis of data from 145 countries. *Tropical Medicine & International Health*, 24(2), 228-238.

Ramesh A, Blanchet K, Ensink JHJ, Roberts B (2015) Evidence on the Effectiveness of Water, Sanitation, and Hygiene (WASH) Interventions on Health Outcomes in Humanitarian Crises: A Systematic Review. *PLOS ONE*, 10(9): e0124688.

Reuters. (2017). Nigeria cholera outbreak kills at least 25 people, prompts emergency measures. Retrieved from https://www.reuters.com/article/us-nigeria-cholera-idUSKBN1C02NV

Robins, S. (2018). The wastefulness of neoliberal water governance: The case of South Africa. *Environment and Planning D: Society and Space*, 36(1), 3-21.

Rothwell, P. M. (2005). Factors that can affect the external validity of randomised controlled trials. *PLoS clinical trials*, 1(1), e9.

Ruppel, O. C., & Mollinga, P. P. (2019). Water, power and the production of neoliberalism in Africa: The case of the Maputo River Basin. *Palgrave Macmillan.*

Savovic, J., Jones, H. E., Altman, D. G., Harris, R. J., Jüni, P., Pildal, J., ... & Schulz, K. F. (2012). Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Annals of internal medicine*, 157(6), 429-438.

Scott, B., Curtis, V., Rabie, T., & Garbrah-Aidoo, N. (2007). Health in our hands, but not in our heads: understanding hygiene motivation in Ghana. *Health Policy and Planning*, 22(4), 225-233.

Shultz, M., & Vaughn, L. (2018). The use of bibliometric analysis for research impact assessment: A comparison of Web of Science, Scopus, and Google Scholar. *Journal of the Medical Library Association*, 106(2).

Singer, P., S, C., & Haushofer, J. (2019). *Are randomized poverty-alleviation experiments ethical?: By Peter Singer, Arthur Baker and Johannes Haushofer*. *Project Syndicate*. Retrieved April 8, 2023, from https://www.project-syndicate.org/commentary/ethics-of-random-controlled-trials-to-fight-poverty-by-peter-singer-et-al-2019-11?barrier=accesspay

Smith, J., & Jones, L. (2019). Why some articles are never cited: A review of the literature. *Journal of Academic Publishing*, 10(3), 123-140.

Sood, A., Jena, A. B., & Lakdawalla, D. N. (2010). The effect of low-baseline risk on the reporting of absolute and relative risk reduction in randomized controlled trials. *Journal of General Internal Medicine*, 25(9), 915-920.

Sultana, F. (2018). Water, technology, and development: Historical dynamics, present realities, and future prospects. *Water Alternatives*, 11(1), 71-91.

Tambo, E., Ai, L., Zhou, X., et al. (2014). Surveillance-response systems: the key to elimination of tropical diseases. *Infection Diseases and Poverty*, 2014;3(1):17.

Thomson, P., & Kateta, N. (2020). Reporting of sex-disaggregated data in evaluations of water, sanitation, and hygiene interventions in low-income settings: A systematic review. *International Journal of Environmental Research and Public Health*, 17(10), 3667.

Thys, S., Mwape, K.E., Lefèvre, P., Dorny, P., Marcotty, T., Phiri, A.M., et al. (2015). Why latrines are not used: communities' perceptions and practices regarding latrines in a Taeniasolium endemic rural area in eastern Zambia. *PLOS Tropical Diseases*, 10;9(9):e0003570.

Tripathy, P., Nair, N., Sinha, R., Rath, S., Gope, R., Rath, S., … Prost, A. (2016). Effect of participatory women's groups and counselling through home visits on children's linear growth in rural eastern India (CARING trial): A cluster-randomised controlled trial. *The Lancet Global Health*, 4(12), e896-e908. https://doi.org/10.1016/S2214-109X(16)30262-0

Tumwebaze, I.K., Mosler, H.-J., & Ndejjo, R. (2018). Effects of interventions on household behaviours related to water, sanitation and hygiene: A systematic review and meta-analysis. *PLOS One*, 13(3), e0194378.

UNICEF. (2014). Guidelines for monitoring and reporting on the situation of children and women. Retrieved from https://data.unicef.org/wp-content/uploads/2015/12/Guidelines-for-Monitoring-and-Reporting-on-the-Situation-of-Children-and-Women.pdf

UNICEF. (2016). Water, sanitation and hygiene (WASH). Retrieved from https://www.unicef.org/wash/

United Nations Water. (n.d.). Retrieved September 18, 2022, from https://www.unwater.org/

UN Water. (n.d.). Decade of Action on Water for Sustainable Development 2018-2028. United Nations. https://www.unwater.org/water-decade/

United Nations. (2008). International Year of Sanitation 2008. Retrieved from https://www.un.org/en/events/sanitationday/background.shtml United Nations. (n.d.). Goal 6: Clean water and sanitation. Retrieved from https://www.un.org/sustainabledevelopment/water-and-sanitation/

United Nations. (2014). Bill and Melinda Gates Foundation pledges $50 million to improve sanitation in India. Retrieved from https://www.un.org/sustainabledevelopment/blog/2014/07/bill-and-melinda-gates-foundation-pledges-50-million-to-improve-sanitation-in-india/

United Nations. (2015). Sustainable Development Goals. Retrieved from https://www.un.org/sustainabledevelopment/sustainable-development-goals/

United Nations General Assembly. (2015). Transforming our world: The 2030 agenda for sustainable development. Retrieved from https://sustainabledevelopment.un.org/post2015/transformingourworld

Van Der Knaap, M., De Bruin, M., & Butler, J. R. A. (2006). Scaling up water and sanitation initiatives: the case of the millennium villages project in sub-Saharan Africa. In Scaling up, scaling down: Overcoming malnutrition in developing countries (pp. 263-286). *Springer*, *Dordrecht*.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.

Viergever, R. F., & Ghersi, D. (2011). The quality of registration of clinical trials. *PLoS One*, 6(2), e14701.

Waddington, H., Snilstveit, B., & White, H. (2009). Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries. *International Initiative for Impact Evaluation (3ie)*.

Ware, M., & Mabe, M. (2015). The STM report: An overview of scientific and scholarly publishing. *International Association of Scientific, Technical and Medical Publishers*.

Whitehead, J., Sheldrick, J., & Patterson, S. (2016). The potential impact of selection bias on intervention effectiveness estimates in randomized trials with restricted eligibility criteria: a simulation study. *BMC Medical Research Methodology*, 16(1), 35.

Wolf, J., Hunter, P. R., Freeman, M. C., Cumming, O., Clasen, T. F., Bartram, J., Higgins, J. P. T., & Johnston, R. B. (2014). Impact of drinking water, sanitation and hand washing with soap on childhood diarrhoeal disease: updated meta-analysis and meta-regression. *Tropical Medicine & International Health*, 19(8), 928–942.

World Bank. (2015). Operationalizing data disaggregation for effective development. Retrieved from https://openknowledge.worldbank.org/handle/10986/22943

World Health Organization. (2014). Handbook for Guideline Development. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/145714/9789241548960_eng.pdf;jsessionid=6CEB86C15D1654D4A6F5E77331E25EBE?sequence=1

World Health Organization. (2018). Hepatitis E outbreak in South Sudan. Retrieved from https://www.who.int/csr/don/30-august-2018-hepatitis-e-south-sudan/en/

World Health Organization. (2019). Progress on household drinking water, sanitation and hygiene 2000-2017: Special focus on inequalities. Retrieved from https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/

World Health Organization (WHO). (2019). Water, sanitation and hygiene for accelerating and sustaining progress on neglected tropical diseases. https://www.who.int/neglected_diseases/WASH_for_NTDs/en/

Yekta, M. M. J. (2021). Scientific zero to one: Some common properties of highly-influential papers. *Malaysian Journal of Library & Information Science*, *26*(3), 1-32.

INFORMATION ON THE STUDIES OF THE DATASET

Authors	Article Title	Year	Source Title	DOI Link
Quattrochi, JP; Coville, A; Mvukiyehe, E; Dohou, CJ; Esu, F; Cohen, B; Bokasola, YL; Croke, K	Effects of a community-driven water, sanitation and hygiene intervention on water and sanitation infrastructure, access, behaviour, and governance: a cluster-randomised controlled trial in rural Democratic Republic of Congo	2021	BMJ GLOBAL HEALTH	http://dx.doi.org/10.1136/bmjgh-2021-005030
Fagerli, K; Gieraltowski, L; Nygren, B; Foote, E; Gaines, J; Oremo, J; Odhiambo, A; Kim, S; Quick, R	Use, Acceptability, Performance, and Health Impact of Hollow Fiber Ultrafilters for Water Treatment in Rural Kenyan Households, 2009-2011	2020	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	http://dx.doi.org/10.4269/ajtmh.19-0862
Solomon, ET; Robele, S; Kloos, H; Mengistie, B	Effect of household water treatment with chlorine on diarrhea among children under the age of five years in rural areas of Dire Dawa, eastern Ethiopia: a cluster randomized controlled trial	2020	INFECTIOUS DISEASES OF POVERTY	http://dx.doi.org/10.1186/s40249-020-00680-9
Swarthout, J; Ram, PK; Arnold, CD; Dentz, HN; Arnold, BF; Kalungu, S; Lin, A; Njenga, SM; Stewart, CP; Colford, JM; Null, C; Pickering, AJ	Effects of Individual and Combined Water, Sanitation, Handwashing, and Nutritional Interventions on Child Respiratory Infections in Rural Kenya: A Cluster-Randomized Controlled Trial	2020	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	http://dx.doi.org/10.4269/ajtmh.19-0779
McQuade, ETR; Platts-Mills, JA; Gratz, J; Zhang, JX; Moulton, LH; Mutasa, K; Majo, FD; Tavengwa, N; Ntozini, R; Prendergast, AJ; Humphrey, JH; Liu, J; Houpt, ER	Impact of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on Enteric Infections in Rural Zimbabwe: The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) Trial	2020	JOURNAL OF INFECTIOUS DISEASES	http://dx.doi.org/10.1093/infdis/jiz179
Gladstone, MJ; Chandna, J; Kandawasvika, G; Ntozini, R; Majo, FD; Tavengwa, NV; Mbuya, MNN; Mangwadu, GT; Chigumira, A; Chasokela, CM; Moulton, LH; Stoltzfus, RJ; Humphrey, JH; Prendergast, AJ	Independent and combined effects of improved water, sanitation, and hygiene (WASH) and improved complementary feeding on early neurodevelopment among children born to HIV-negative mothers in rural Zimbabwe: Substudy of a cluster-randomized trial	2020	PLOS MEDICINE	http://dx.doi.org/10.1371/journal.pmed.1002766
Humphrey, JH; Mbuya, MNN; Ntozini, R; Moulton, LH; Stoltzfus, RJ; Tavengwa, NV; Mutasa, K; Majo, F; Mutasa, B; Mangwadu, G; Chasokela, CM; Chigumira, A;	Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial	2019	LANCET GLOBAL HEALTH	http://dx.doi.org/10.1016/S2214-109X(18)30374-7

Chasekwa, B; Smith, LE; Tielsch, JM; Jones, AD; Manges, AR; Maluccio, JA; Prendergast, AJ				
Stewart, CP; Dewey, KG; Lin, A; Pickering, AJ; Byrd, KA; Jannat, K; Ali, S; Rao, G; Dentz, HN; Kiprotich, M; Arnold, CD; Arnold, BF; Allen, LH; Shahab-Ferdows, S; Ercumen, A; Grembi, JA; Naser, A; Rahman, M; Unicomb, L; Colford, JM; Luby, SP; Null, C	Effects of lipid-based nutrient supplements and infant and young child feeding counseling with or without improved water, sanitation, and hygiene (WASH) on anemia and micronutrient status: results from 2 cluster-randomized trials in Kenya and Bangladesh	2019	AMERICAN JOURNAL OF CLINICAL NUTRITION	http://dx.doi.org/10.1093/ajcn/nqy239
Null, C; Stewart, CP; Pickering, AJ; Dentz, HN; Arnold, BF; Arnold, CD; Benjamin-Chung, J; Clasen, T; Dewey, KG; Fernald, LCH; Hubbard, AE; Kariger, P; Lin, A; Luby, SP; Mertens, A; Njenga, SM; Nyambane, G; Ram, PK; Colford, JM	Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial	2018	LANCET GLOBAL HEALTH	http://dx.doi.org/10.1016/S2214-109X(18)30005-6
Morris, JF; Murphy, J; Fagerli, K; Schneeberger, C; Jaron, P; Moke, F; Juma, J; Ochieng, JB; Omore, R; Roellig, D; Xiao, LH; Priest, JW; Narayanan, J; Montgomery, JM; Hill, V; Mintz, E; Ayers, TL; O'Reilly, CE	A Randomized Controlled Trial to Assess the Impact of Ceramic Water Filters on Prevention of Diarrhea and Cryptosporidiosis in Infants and Young Children-Western Kenya, 2013	2018	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	http://dx.doi.org/10.4269/ajtmh.17-0731
Humphrey, JH; Jones, AD; Manges, A; Mangwadu, G; Maluccio, JA; Mbuya, MNN; Moulton, LH; Ntozini, R; Prendergast, AJ; Stoltzfus, RJ; Tielsch, JM; Chasokela, C; Chigumira, A; Heylar, W; Hwena, P; Kembo, G; Majo, FD; Mutasa, B; Mutasa, K; Rambanepasi, P; Sauramba, V; Tavengwa, NV; Van Der Keilen, F; Zambezi, C; Chidhanguro, D; Chigodora, D; Chipanga, JF; Gerema, G; Magara, T; Mandava, M; Mavhudzi, T; Mazhanga, C; Muzaradope, G; Mwapaura, MT; Phiri, S; Tengende, A	The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) Trial: Rationale, Design, and Methods	2018	CLINICAL INFECTIOUS DISEASES	http://dx.doi.org/10.1093/cid/civ844

Pickering, AJ; Djebbari, H; Lopez, C; Coulibaly, M; Alzua, ML	Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial	2018	LANCET GLOBAL HEALTH	http://dx.doi.org/10.1016/S2214-109X(15)00144-8
Cha, S; Kang, D; Tuffuor, B; Lee, G; Cho, J; Chung, J; Kim, M; Lee, H; Lee, J; Oh, C	The Effect of Improved Water Supply on Diarrhea Prevalence of Children under Five in the Volta Region of Ghana: A Cluster-Randomized Controlled Trial	2015	INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	http://dx.doi.org/10.3390/ijerph121012127
Christensen, G; Dentz, HN; Pickering, AJ; Bourdier, T; Arnold, BF; Colford, JM; Null, C	Pilot Cluster Randomized Controlled Trials to Evaluate Adoption of Water, Sanitation, and Hygiene Interventions and Their Combination in Rural Western Kenya	2015	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	http://dx.doi.org/10.4269/ajtmh.14-0138
Luoto, J; Levine, D; Albert, J; Luby, S	Nudging to use: Achieving safe water behaviors in Kenya and Bangladesh	2015	JOURNAL OF DEVELOPMENT ECONOMICS	http://dx.doi.org/10.1016/j.jdeveco.2014.02.010
Rosa, G; Majorin, F; Boisson, S; Barstow, C; Johnson, M; Kirby, M; Ngabo, F; Thomas, E; Clasen, T	Assessing the Impact of Water Filters and Improved Cook Stoves on Drinking Water Quality and Household Air Pollution: A Randomised Controlled Trial in Rwanda	2014	PLOS ONE	http://dx.doi.org/10.1371/journal.pone.0091011
Mengistie, B; Berhane, Y; Worku, A	Household Water Chlorination Reduces Incidence of Diarrhea among Under-Five Children in Rural Ethiopia: A Cluster Randomized Controlled Trial	2014	PLOS ONE	http://dx.doi.org/10.1371/journal.pone.0077887
Gunther, I; Schipper, Y	PUMPS, GERMS AND STORAGE: THE IMPACT OF IMPROVED WATER CONTAINERS ON WATER QUALITY AND HEALTH	2013	HEALTH ECONOMICS	http://dx.doi.org/10.1002/hec.2852
Stauber, CE; Kominek, B; Liang, KR; Osman, MK; Sobsey, MD	Evaluation of the Impact of the Plastic BioSand Filter on Health and Drinking Water Quality in Rural Tamale, Ghana	2013	INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	http://dx.doi.org/10.3390/ijerph9113806
Patel, MK; Harris, JR; Juliao, P; Nygren, B; Were, V; Kola, S; Sadumah, I; Faith, SH; Otieno, R; Obure, A; Hoekstra, RM; Quick, R	Impact of a Hygiene Curriculum and the Installation of Simple Handwashing and Drinking Water Stations in Rural Kenyan Primary Schools on Student Health and Hygiene Practices	2012	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	http://dx.doi.org/10.4269/ajtmh.2012.11-0494
du Preez, M; Conroy, RM; Ligondo, S; Hennessy, J; Elmore-Meegan, M; Soita, A; McGuigan, KG	Randomized Intervention Study of Solar Disinfection of Drinking Water in the Prevention of Dysentery in Kenyan Children Aged under 5 Years	2012	ENVIRONMENTAL SCIENCE & TECHNOLOGY	http://dx.doi.org/10.1021/es2018835
Boisson, S; Kiyombo, M; Sthreshley, L; Tumba, S; Makambo, J; Clasen, T	Field Assessment of a Novel Household-Based Water Filtration Device: A Randomised, Placebo-Controlled Trial in the Democratic Republic of Congo	2011	PLOS ONE	http://dx.doi.org/10.1371/journal.pone.0012613
Albert, J; Luoto, J; Levine, D	End-User Preferences for and Performance of Competing POU Water Treatment Technologies among the Rural Poor of Kenya	2011	ENVIRONMENTAL SCIENCE & TECHNOLOGY	http://dx.doi.org/10.1021/es100566
Boisson, S; Schmidt, WP; Berhanu, T; Gezahegn, H; Clasen, T	Randomized Controlled Trial in Rural Ethiopia to Assess a Portable Water Treatment Device	2010	ENVIRONMENTAL SCIENCE & TECHNOLOGY	http://dx.doi.org/10.1021/es900664

du Preez, M; Conroy, RM; Wright, JA; Moyo, S; Potgieter, N; Gundry, SW	Short Report: Use of Ceramic Water Filtration in the Prevention of Diarrheal Disease: A Randomized Controlled Trial in Rural South Africa and Zimbabwe	2010	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	http://dx.doi.org/10.4269/ajtmh.2008.79.696
Garrett, V; Ogutu, P; Mabonga, P; Ombeki, S; Mwaki, A; Aluoch, G; Phelan, M; Quick, RE	Diarrhoea prevention in a high-risk rural Kenyan population through point-of-use chlorination, safe water storage, sanitation, and rainwater harvesting	2009	EPIDEMIOLOGY AND INFECTION	http://dx.doi.org/10.1017/S095026880700026X
Crump, JA; Otieno, PO; Slutsker, L; Keswick, BH; Rosen, DH; Hoekstra, RM; Vulule, JM; Luby, SP	Household based treatment of drinking water with flocculant-disinfectant for preventing diarrhoea in areas with turbid source water in rural western Kenya: cluster randomised controlled trial	2008	BMJ-BRITISH MEDICAL JOURNAL	http://dx.doi.org/10.1136/bmj.38512.618681.E0
		2008		
		2005		