HUMANITARIAN EVIDENCE PROGRAMME

Impact of WASH Interventions during Disease Outbreaks in Humanitarian Emergencies: A systematic review protocol

UKaid from the British people

Tufts University School of Engineering

Tufts University Friedman School of Nutrition Science and Policy

Feinstein International Center

OXFAM

Authors
Travis Yates, Tufts University
Jelena Vijcic, Independent Research Scientist
Dr Myriam Leandre Joseph, Physician/Consultant
Dr Daniele Lantagne, Assistant Professor, Tufts University.

Contact
Dr Daniele Lantagne, Tufts University, Medford, MA: daniele.lantagne@tufts.edu.

Funding
This is a report commissioned by the Humanitarian Evidence Programme, a partnership between Oxfam and Feinstein International Center at Tufts University, and funded by the Department for International Development. This material has been funded by UK aid from the UK Government, however, the views expressed do not necessarily reflect the UK Government's official policies.

Picture
As part of Oxfam's cholera response in Juba, teams of public health volunteers have been teaching affected communities about the importance of keeping themselves and their environment clean. May 2014. Kieran Doherty/Oxfam.

CONTENTS

ACRONYMS

CASP	Critical Appraisal Skills Programme
CDC	Centre for Disease Control
CENTRAL	Cochrane Centre Registers for Clinical Trials
CINAHL	Cumulative Index to Nursing and Allied Health
CSB	Corn-soy blend
CRD	Centre for Reviews and Dissemination
CMAM	Community-based management of acute malnutrition
DFID	Department for International Development
ENN	Emergency Nutrition Network
ELRHA	Enhanced Learning and Research for Humanitarian Assistance
EMBASE	Excerpta Medica Database
FANTA	Food and Nutrition Technical Assistance
FAO	Food and Agriculture Organisation
HPG	Humanitarian Policy Group
INGOs	International non-governmental organisations
IMEMR	Index Medicus for Eastern Mediterranean Region
IMSEAR	Index Medicus for South-East Asian Region
LILACS	Latin America Caribbean Health Sciences Literature
MAM	Moderate acute malnutrition
MSF	Medicine Sans Frontiers
MUAC	Mid-Upper Arm Circumference
ODI	Overseas Development Institute
NCHS	National Center for Health Statistics
SAM	Severe acute malnutrition
SC	Supper Cereal
RCT	Randomised controlled trials
RUTF	Ready-to-use therapeutic foods
RUF	Ready-to-use foods
RUSF	Ready-to-use supplementary foods
R4D	Research for Development
TSF	Target supplementary feeding
USA	United States of America
UK	United Kingdom
UNHCR	United Nations High Commissioner for Refugees
UNSCN	United Nations Standing Committee on Nutrition
USAID	United States Agency for International Development
UNICEF	United Nations Children's Fund
WHO	World Health Organization

1 BACKGROUND

1.1 DESCRIPTION OF THE PROBLEM

According to the World Health Organization (WHO):

> *A disease outbreak is the occurrence of cases of disease in excess of what would normally be expected in a defined community, geographical area or season. An outbreak may occur in a restricted geographical area, or may extend over several countries. It may last for a few days or weeks, or for several years. A single case of a communicable disease long absent from a population, or caused by an agent (e.g. bacterium or virus) not previously recognized in that community or area, or the emergence of a previously unknown disease, may also constitute an outbreak and should be reported and investigated [1].*

Thus, an outbreak could be defined as an increase above the normal background rate of malaria, or defined as one case of Ebola in a country where the virus had not previously been recognized. Worldwide, the number and diversity of disease outbreaks has increased from 1980-2013 [2]. During those 34 years, 12,102 outbreaks of 215 human infectious diseases, comprising more than 44 million cases, occurred in 219 nations. The most common human specific outbreaks during this time period were: adenovirus, cholera, enterovirus, gastroenteritis, hepatitis B, legionellosis, malaria, measles, meningitis, mumps, pertussis, rotavirus infection, rubella, and typhoid. The most common zoonotic outbreaks were anthrax, camplylobacterosis, chikungunya, cryptosporidiosis, dengue, *E. coli* diarrhea, hepatitis A, hepatitis E, influenza A, salmonellosis, shigellosis, trichinosis, and tuberculosis. Although the number of outbreaks increased with time in the human population both in total number and richness of causal diseases, outbreak cases *per capita* appear to be declining over time, indicating global improvements in prevention, early detection, control and treatment are becoming more effective at reducing the number of people infected.

WASH interventions as an outbreak response strategy

Outbreak response strategies vary depending on the disease type, resources, and local context. Interventions cover a variety of sectors including: medical, public health, and/or engineering aspects. Preventative vaccines, oral rehydration solution (ORS), and medicines are some common health focused outbreak strategies. Water, sanitation, and hygiene (WASH) interventions are other outbreak mitigation strategies that aim to prevent and control waterborne and communicable diseases [3], [4]. WASH interventions are critical to the prevention and control of outbreaks, as:

- WASH coverage and provision *prevents* outbreaks caused by waterborne disease agents, such as *E. coli,* cholera, and gastroenteritis [5].
- WASH interventions can assist providers and responders in *controlling* the spread and transmission of disease, both in treatment facilities and in communities.

Providing safe water and promoting handwashing are common WASH interventions in outbreaks, but interventions could also include managing the local environmental hazards like rubbish disposal or increasing latrine use. Infectious disease outbreaks that are not necessarily waterborne (i.e. Ebola) can also benefit from WASH interventions by promoting hand and environmental hygiene. Emergency WASH interventions, as in response to an outbreak, are usually not initially intended to provide long-term sustainable programming, but instead provide rapid relief to minimize the impact or spread of an outbreak [3].

Organizational response

As a response to any type of disaster or emergency, 'humanitarian aid' includes a range of interventions intended to save lives and alleviate human suffering. The United Nations (UN) has an agency dedicated to coordinating the humanitarian response effort, the UN Office for the Coordination of Humanitarian Affairs (OCHA). In 2014, the OCHA funding appeal was $17.9 billion (USD) and intended to support more than 50 million people in 31 countries [6].

Figure 1: Cholera treatment center

WHO/Paul Garwood

Local and national governments in low- and middle-income countries (LMIC) are often unable to effectively respond to disease outbreaks. In this case, the WHO has resources to help local governments and protect the general population. The WHO typically leads the UN or OCHA response in an outbreak, but requires significant coordination with the local government, as well as other UN agencies. For instance, the UN Children's Fund (Unicef) guides WASH interventions and the United Nations High Commission for Refugees (UNHCR) is the technical lead in refugee settings. Non-governmental organizations (NGOs) (e.g. Medecins Sans Frontieres (MSF), the Red Cross and Red Crescent Societies (ICRC), or International Medical Corps (IMC)) specialize in outbreak/emergency response and regularly manage hospitals or clinics. NGOs or UN agencies also set up specialized treatment centers for some outbreaks, as in the case of a cholera treatment centers in Haiti or Ebola treatment units in Sierra Leone. Additionally, the Global Outbreak Alert and Response Network (GOARN) is a network of organizations with resources and expertise to rapidly respond to outbreaks in conjunction with the WHO. The Centers for Disease Control and Prevention (CDC) also has extensive expertise in outbreak management and coordinates with the WHO, governmental, NGO, and local partners. All of these partners are in constant communication and coordination with national governments who will eventually take over more responsibility and transition out of the emergency.

Outbreaks and emergencies

Disease outbreaks can be a primary (direct) cause of an emergency, but they can also spring up after other emergencies as secondary (indirect) emergencies. An example of a primary disease outbreak would be the 1994 cholera outbreak in Congo where mortality rates rose 20-30 times above the baseline rates, 50,000 Rwandan refugees died from cholera over a four week period [7] -[8]. More recently, hepatitis E outbreaks have occurred in several African refugee camps and have killed hundreds of people, especially pregnant women and children[9]. Secondary emergencies have been occurred specifically after flooding emergencies and emergencies that cause large population displacement resulting in significant increases in waterborne disease risk [10]–[15].

Scope

In this review, we will investigate the impact of eight WASH interventions in preventing (reducing the risk of) and controlling outbreaks in LMIC, with particular focus on three diseases of current concern to the response community – cholera, Ebola, and Hepatitis E. Additionally, we will explore economic outcomes related to WASH interventions within an outbreak.

This investigation of WASH interventions in disease outbreaks will be a systematic review of both published and grey literature. Grey literature is excluded in most systematic reviews; however, in the humanitarian sector, grey literature (e.g. NGO or UN reports) provides a valuable source of information that is often not published in academic journals. We aim to differentiate evidence on WASH interventions in outbreaks by disease type, population type (i.e. refugee, internally displaced persons (IDPs), community), geographic region, urban/rural

setting, training components, concurrent emergencies, complimentary interventions, gender, age, impact, and other characteristics.

1.2 WHY IT IS IMPORTANT TO DO THIS REVIEW

The impact and effects of WASH interventions in development contexts has been extensively studied and debated [16]–[19]. In contrast, there is a general 'lack of evidence' in emergency response interventions [20]. The evidence on WASH interventions in emergency response situations in general – and in particular outbreaks of Ebola, cholera, and Hepatitis E – have not yet been systematically reviewed; however, WASH interventions are undertaken in the vast majority of outbreak responses to prevent and control the disease burden. The lack of research is often attributed to a limited staff capacity, the priority of need for immediate response, ethical considerations, and access. Also, WASH interventions that are intended to prevent or reduce disease transmission may have difficulty showing impact because of the uncertainty of knowing the 'future' or 'potential' disease rate unless there is a rigorous study design that is not often conducted in an emergency due to ethical considerations of having a control group in emergencies. And many emergency response interventions have been extrapolated from development contexts, leading to an insufficient understanding of what would be an appropriate response [21], [22]. Research has also shown that many emergency responders default to familiar interventions previously used, which may not be the most effective response [23], [24]. In emergencies like outbreaks, 'intuition' and 'if it worked before it will work again' are mentalities of relief professionals faced with complex situations with unknown consequences [24], [25]. This implies that some interventions may be used in inappropriate contexts, and studies have shown that tater treatment strategies have failed when used in contexts that are outside the recommended context [24], [26]. The effectiveness of interventions is a function of physical parameters, but also social factors, such as community acceptance and ease of use [10], [21], [24].

There has been work recently completed by the London School of Hygiene and Tropical Medicine (LSHTM) looking specifically at published literature on WASH interventions for cholera-response [27]. However, this work did not consider unpublished (grey literature) from UN agencies or NGOs and it did not consider lessons that could be adapted from other outbreaks. Additionally, there have been literature reviews of individual WASH interventions in the past (such as household water treatment) [28], but there has been no systematic review including all WASH interventions in outbreaks that incorporates information from grey literature to complete a cohesive picture of all WASH interventions in response to outbreaks. This work aims to fill this gap.

1.3 DESCRIPTION OF THE INTERVENTIONS

For the purpose of this review, we have separated WASH interventions into eight specific outbreak response intervention categories. These interventions were selected based on the scope of interventions that are most commonly implemented in response to outbreaks in LMIC. These interventions can be implemented along-side or in combination with each other; however, all aim to prevent and control the disease burden during disease outbreaks. Interventions can also be implemented in conjunction with health interventions (i.e. vaccines or other treatment). The eight WASH interventions included in this review are: 1) increasing water access; 2) source-based water treatment; 3) household water treatment (HWT); 4) promotion of hand hygiene at critical times; 5) distribution of soap and/or hygiene kits; 6) environmental hygiene; 7) installation of temporary or permanent sanitation facility; and 8) distribution of latrine alternatives. These interventions are separated into interventions to assist in the search strategy and identify influential factors in the causal chain described in section 1.4 below; a different arrangement of interventions may be presented in the final report.

1) Increasing water access

Figure 2: Water tankering in Syria (ICRC, 2015)

Access to water is critical for outbreak-affected populations. Existing water sources can be damaged or contaminated with a waterborne disease, or overwhelmed by a sudden influx of displaced persons. Increasing water access is a necessary step in providing potable water, but also enables hygiene and sanitation practices. In outbreaks, there is rarely time or focus for new construction of water points. Thus the most common water access interventions are: 1) repair/cleaning existing sources; and 2) water tankering. *Repairing or cleaning existing sources*, like wells or springs, are often one-time interventions that restore water sources familiar to the local populations. *Water tankering* (Figure 2) hauls water from another source, bringing it to the outbreak-affected population.

2) Source-based water treatment options

Figure 3: Chlorine Dispenser used in the DRC (Armitage, 2013)

In contexts where water access is secured, source-based water treatment aims to improve water quality during collection. Most source-based treatments use chlorine solution or chlorine tablets to treat water because it effectively protects against most waterborne diseases [29]. Source-based treatment interventions are differentiated by the chlorine delivery method and beneficiary involvement. *Bucket chlorination* is when a dedicated staff member is stationed by the water source and adds a dose of chlorine directly into the recipient's water collection container. *Chlorine Dispensers* are hardware installed next to a water source; recipients collect water, and then turn the Dispenser valve to dose their own container (Figure 3). *Pot chlorination* is hardware installed in a well, intended to slowly disperse chlorine over an extended time; the beneficiary is not involved. *Temporary pumping and storage of surface water* is the pumping of river or lake water into large bladders or tanks, and then sometimes treated with a flocculent that helps to settle suspended solids, and dosed with chlorine; beneficiaries are not involved.

3) Household water treatment technologies

Figure 4: Beneficiary with PuR Purifier of Water sachets (World Vision, 2013)

Household water treatment (HWT) interventions are another WASH intervention used in contexts where water access is secured but water quality is not adequate. HWT interventions are differentiated by consumable and durable treatments. Consumable items include *flocculent/disinfectant packets*, like P&G® Purifier of Water (Figure 4), or *chlorine tablets* like Aquatabs that are distributed to households to dose specific volumes of water typical for a household (20-25L). Durable HWT include water filters such as: *hollow fiber filters* like LifeStraw® *or filter systems* with ceramic elements are often used. Solar disinfection, SODIS, is another HWT technology that uses ultraviolet radiation and heat to disinfect household drinking water. Finally, boiling is sometimes promoted in emergency situations.

4) Hygiene promotion

Figure 5: Hygiene education in schools (Global Giving)

Personal hygiene during outbreaks is important to prevent the spread of disease. Hygiene promotion is used to educate outbreak-affected populations on the specific disease and mitigation strategies. Often in outbreaks, hygiene promotion is condensed to key messages, such as hand-washing at critical times. Promotion can be at schools, large community groups, or at the household level (Figure 5). Handwashing promotion may also include handwashing stations or tippy taps installed near latrines, homes or schools.

In recent years, there has been a sanitation strategy that focuses on hygiene education and community involvement to address the problem of open defecation. Community Led Total Sanitation (CLTS) has been widely promoted, mostly in development settings, to encourage communities to build their own latrines from locally available materials; specifically, no materials are given to the population. Education through an outside facilitator is intended to influence the population to want to be open defecation free (ODF) and find their own solutions to address open defecation. Similarly, Community Approaches to Total Sanitation (CATS) also encourages social change and ODF communities; however, some assistance could be given in the form of materials or cash to help build latrines. Both CLTS and CATS are highly dependent on hygiene promotion to inform communities to the hazards of open defecation; thus, for this review, we will consider both interventions as hygiene promotion.

5) Distribution of soap and/or hygiene kits

Hygiene-kit distributions provide outbreak-affected populations with soap, buckets, feminine hygiene materials, toothbrushes, and other materials necessary to reduce the risk of disease transmission. Hygiene kits can be distributed as standalone packages, or a component of a large distribution of non-food items (NFIs) or core relief items (CRIs) that includes materials such as bednets, cooking pots, or other materials. An alternative to providing physical materials, subsidies, vouchers, and cash transfers offer flexibility to the disaster-affected households. These options enable the households to prioritize their own needs, but require access to markets.

6) Environmental hygiene interventions

The local environment (household, school, market) is often a route of disease transmission, and in many outbreaks, there are local conditions that increase environmental hazards. Environmental hygiene efforts aim to protect populations from existing or new risks by reducing environmental pathways of disease. Two examples of environmental hygiene interventions are rubbish collection and household spraying. *Rubbish collection* is the removal, management, and disposal of rubbish, often most needed in a refugee camp or informal settlements to minimize vectors that spread disease, like flies and rats. *Household spraying* is when a team of people sanitize a home or building that has potential for risk for contamination; for example, a strong chlorine solution is used to sanitize an Ebola patient's home.

7) Installation of temporary or permanent sanitation facilities

Figure 6: Latrine construction in a refugee camp (IMC)

Management of fecal waste is fundamental to minimize the spread of fecal-oral diseases. Human feces are a primary transmission route of many waterborne diseases. Proper management of both waste and disease vectors are necessary. In most outbreak response situations, sanitation facility interventions are the installation of permanent or temporary latrines. Sanitation facility is a general term, typically referring to a latrine (Figure 6). Construction of a *permanent latrine* may be with a concrete pad or a strong structure that is intended to last for several years. *Temporary latrines*, like transportable port-o-johns or plastic tarpaulin, can also be also used.

8) Distribution and management of latrine alternatives

In some contexts, formal sanitation facilities my not be a viable because of space, timing, or water table constraints. There is a significant amount of innovation in this space. One innovation is the distribution of bags to households intended for single use human waste needs (i.e. pee-poo bags).

Combination and synergies

In many contexts, several interventions described above could be implemented together or with other sectors like health or nutrition. Following an emergency situation, the needs of emergency-affected populations are usually in excess of what a single intervention can address, thus it is common for one or more agencies to implement several interventions in combination. In some situations, WASH interventions are seen as package that addresses water, sanitation, and hygiene needs of a population. With interventions being carried out in unison, there can be synergies that have an additive or diminished effect. We will separate interventions where possible, but also acknowledge the synergies when separation cannot be achieved.

Non-health related interventions

There are many non-health related interventions that address the safety and well-being of disaster affected populations. This can be described as 'quality of life' aspects that are often expressed as protection (i.e. feeling 'safer') or some form of equality (i.e. being less marginalized or stigmatized). For example, women may report feeling safer and less stigmatized when they have Menstrual Health Management (MHM) materials and a latrine nearby. Quality of life impacts are important for this review; however, will be only considered as a result of the interventions listed above.

1.4 HOW THE INTERVENTION MIGHT WORK

To evaluate WASH interventions in disease outbreaks we will follow the assessment principles by Howard White (2009) including: 1) map out the intervention causal chain; 2) understand the context; 3) anticipate heterogeneity; 4) rigorous evaluation of impact using credible counterfactual evidence; 5) rigorous factual analysis; and 6) use of mixed methods [30]. For our systematic review, we use each of the six principles below to assess the eight WASH interventions.

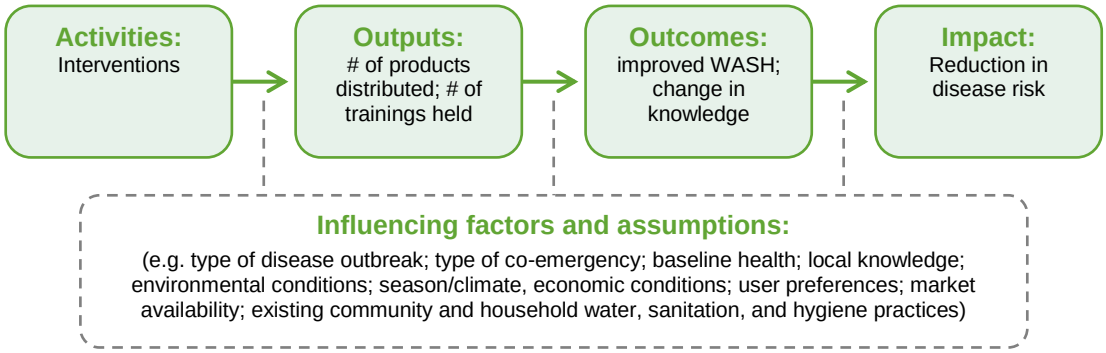
Mapping the causal chain

Beginning with the framework that outbreak-affected populations are at an increased risk of disease, the theory of change that underpins all WASH interventions in outbreaks is:

WASH interventions can reduce the increased risk of disease until such time as the outbreak has ended.

A theory of change will be incorporated into the review by analyzing the outcomes and impacts that lead to risk reduction from WASH interventions in the context of a disease outbreak, and clarifying underlying assumptions. The logic model is a framework that transitions between intervention activities that eventually lead to community impact (Figure 7).

Figure 7: Theory of change



Source: Authors

Activities of WASH interventions during outbreaks generally fall into one of two categories: 1) the distribution of products (i.e. soap, chlorine tablets); or 2) provision of services (i.e. well chlorination, handwashing promotion). Products and services can be provided with, or without, community involvement or training (i.e. nonfood item distributions compared to programs focused on community health workers reaching a wide population).

At this point, we are unsure of the completeness and robustness of the studies that will be included in this review; however, we have a quality assurance process (Section 3 and 4) and will highlight any gaps in programming activities.

Outputs of WASH interventions are generally reported as the number of products delivered or services completed by the implementing agency; for example: the number of buckets distributed or the number families that attended a handwashing seminar.

Outcomes are the direct result of the intervention on the population; for example: use of the distributed product or service to improve drinking water quality, increased knowledge, or a reduced exposure to contamination.

Impacts show the final result of an intervention. For WASH interventions in outbreaks, the impacts are the prevention and control of disease transmission; this is often shown as a reduction in disease prevalence or incidence or a reduction in mortality. Impact can be

difficult to assess with interventions intended to prevent disease during an outbreak because of the uncertainty of potential future cases.

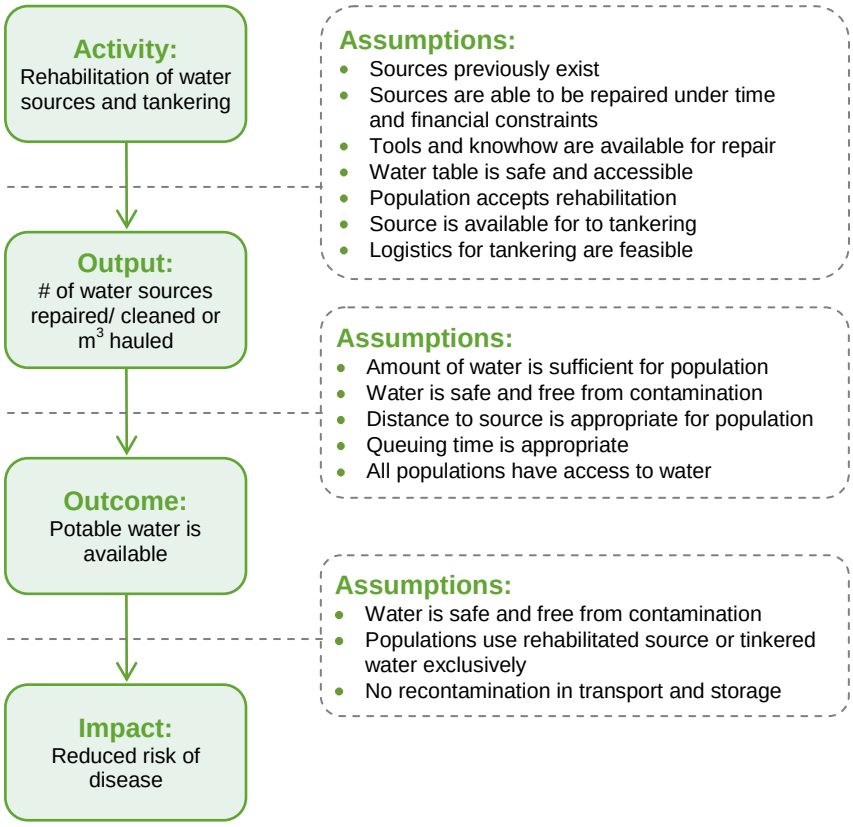
The wide variety of WASH interventions creates a complex causal chain that is difficult to analyze in sufficient detail as one intervention. For example, the activities and outcomes for a behavior change intervention, such as handwashing, is quite different than provision of services, such as a building a latrine or treating water. In order to properly assess activities and assumptions, we have developed a separate causal chain for each of the eight hygiene interventions.

In keeping with the Theory Based Impact Evaluation by Howard White (2009), the causal chain is presented as separate interventions, but the remaining five criteria are presented together. We feel this is appropriate because while there are differences in interventions, the situation in which they are assessed and ability to be broadly applied is common among all the interventions.

1) Increasing water access

The causal chain for the rehabilitation or cleaning of water sources relies on the feasibility and availability to repair damaged sources or clean contaminated sources. With the existing infrastructure, populations are likely familiar with the operation and use of the water source. Thus, critical assumptions are that the source can be repaired or cleaned, and that it provides an adequate amount of water for the population for drinking, as well as, sanitation and hygiene needs. Water tankering is another intervention that increases water access. Critical assumptions for water tankering are that a source is available to collect water in a timely manner with road access for hauling.

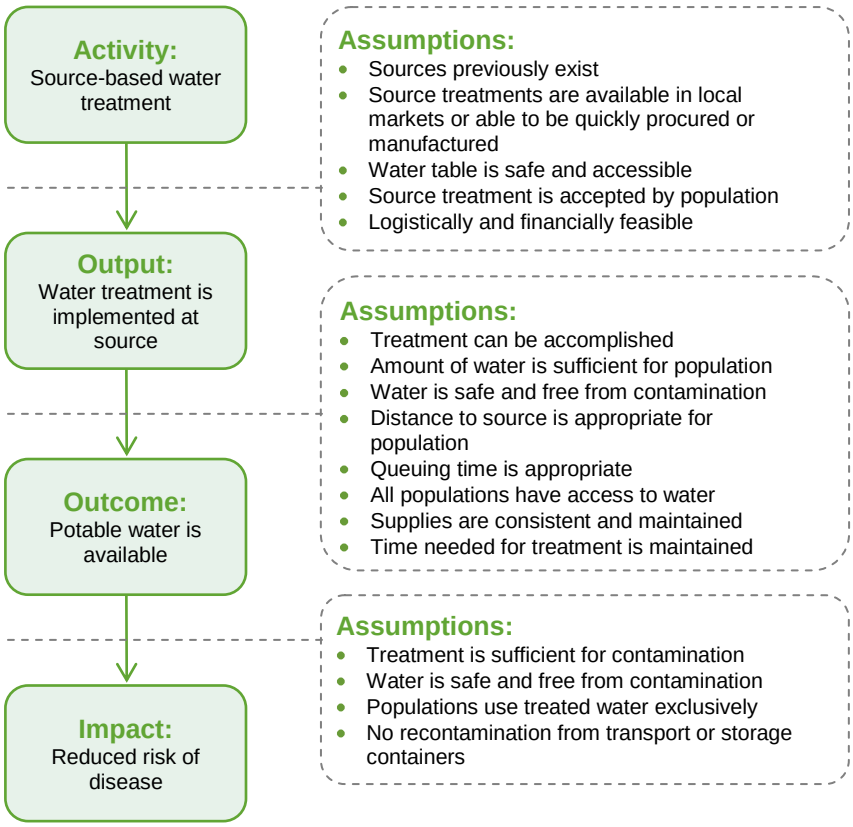
Figure 8: Rehabilitation of water sources program theory



2) Source-based water treatment options

The program theory for source-based water treatment is that a sufficient amount of water quantity is accessible, but water quality is lacking at point sources (e.g. protected wells or springs) and surface water. The critical assumption is that access to the treatment is available at all the sources and at all the times the population collects water. Source based treatment, like Dispensers, may be a new treatment method for the population and require education on correct use.

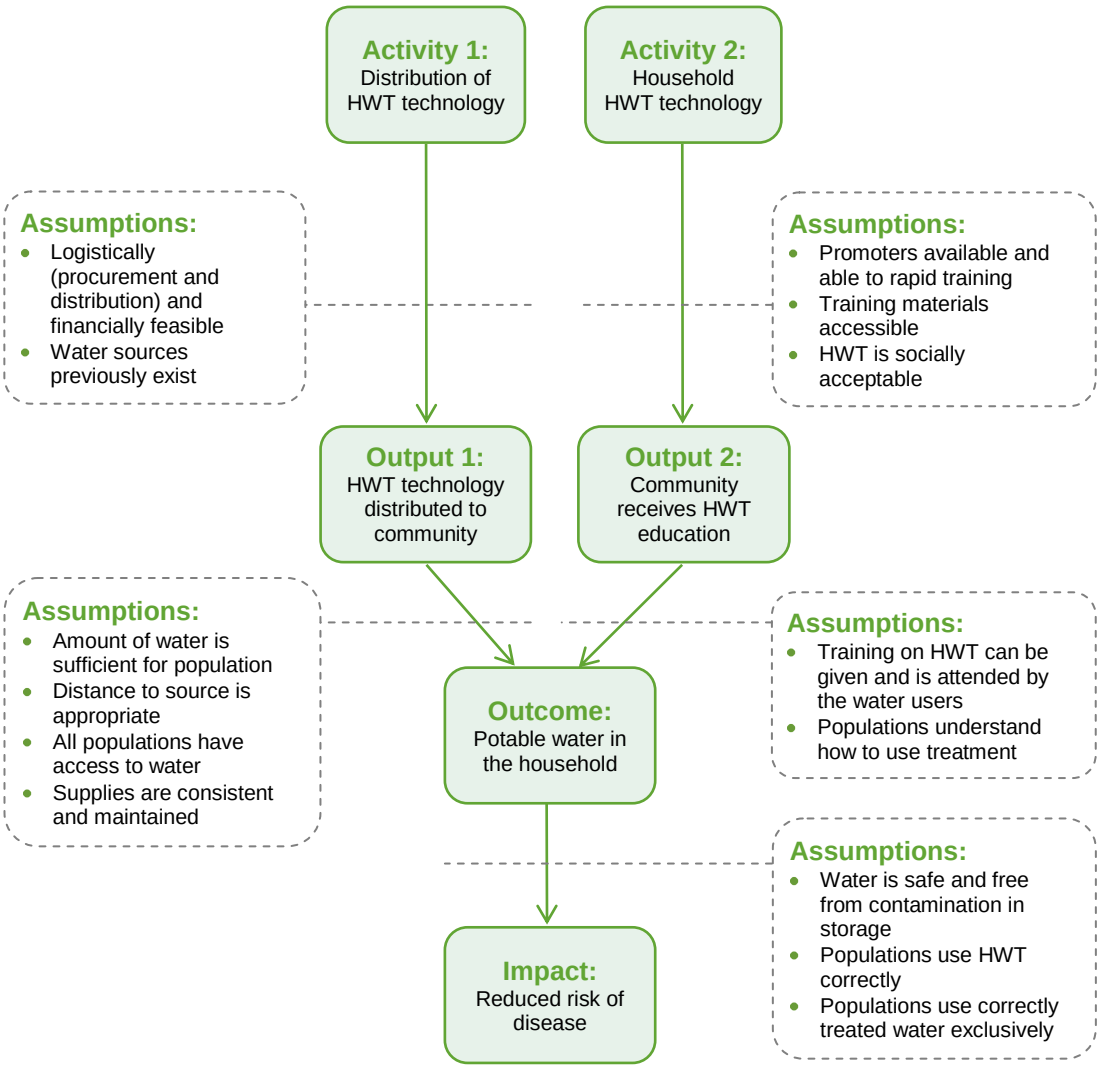
Figure 9: Source-based water treatment program theory



3) Household water treatment technologies

Household water treatment (HWT) program theory is based on adequate access to some water supply that is then treated at the home. This requires the one-time or continued distribution of treatment materials and also an understanding of how to use the treatment method. The critical assumptions are that the treatment is appropriate for the water conditions, households know how to use the treatment correctly, households use the treatment every day, and are able to acquire materials needed for ongoing use.

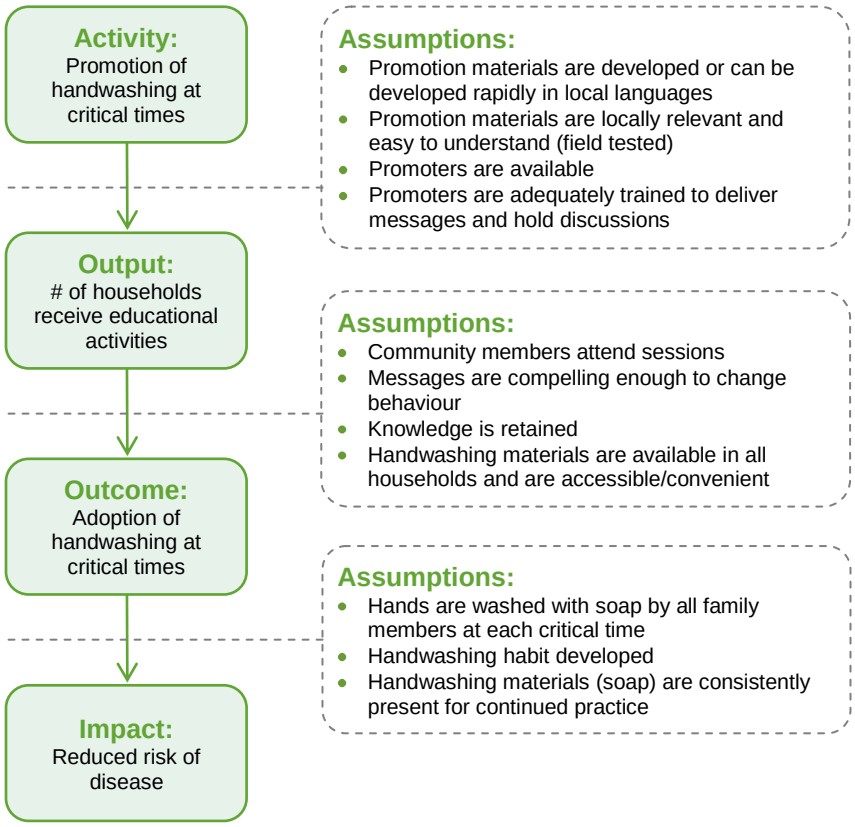
Figure 10: Household water treatment program theory



4) Hygiene promotion

The program theory for hand hygiene at critical time is dependent on breaking the fecal-oral route of contamination. The critical assumptions are that populations have access to soap or ash and populations quickly adopt hygiene messages, including latrine use in CLTS or CATS interventions.

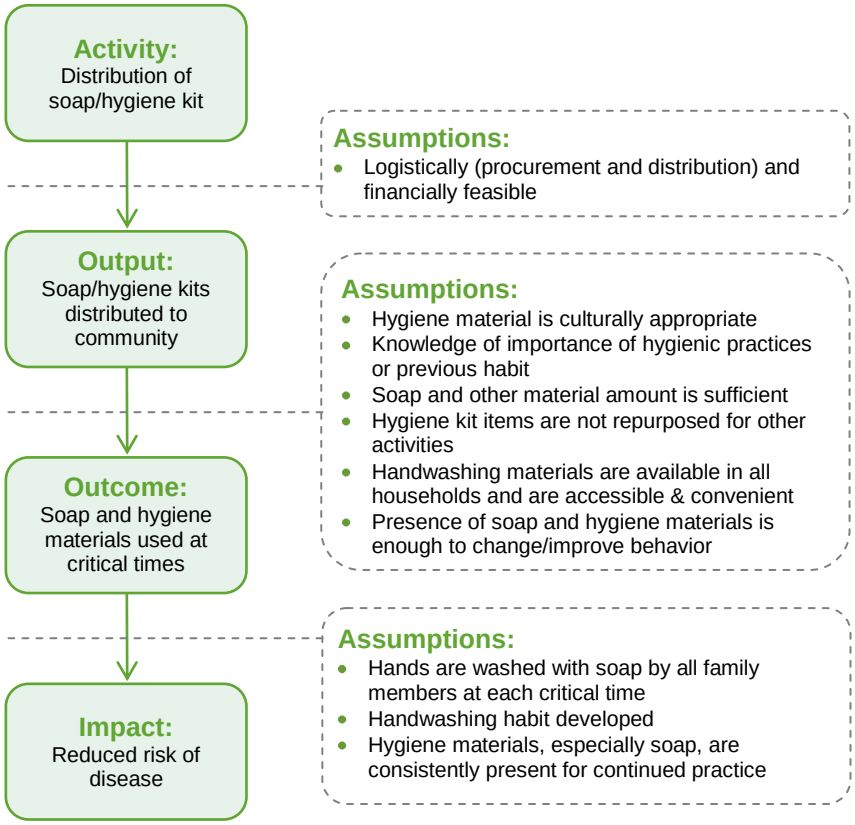
Figure 11: Promotion of hand hygiene at critical times program theory



5) Distribution of soap and/or hygiene kits

The program theory for the distribution of soap and/or hygiene kits is that materials are distributed directly to outbreak-affected populations to reduce their risk of transmission. The critical assumption is that populations already know how to correctly use or is able to quickly learn correct use of items in the kit, because distributions typically have concurrent or no training components. Maintaining consistent supplies to households of different sizes or households with small children is also a challenge. With cash or vouchers, there are assumptions that hygiene materials can be acquired in the markets and prioritized by beneficiary, as opposed to food or other needs.

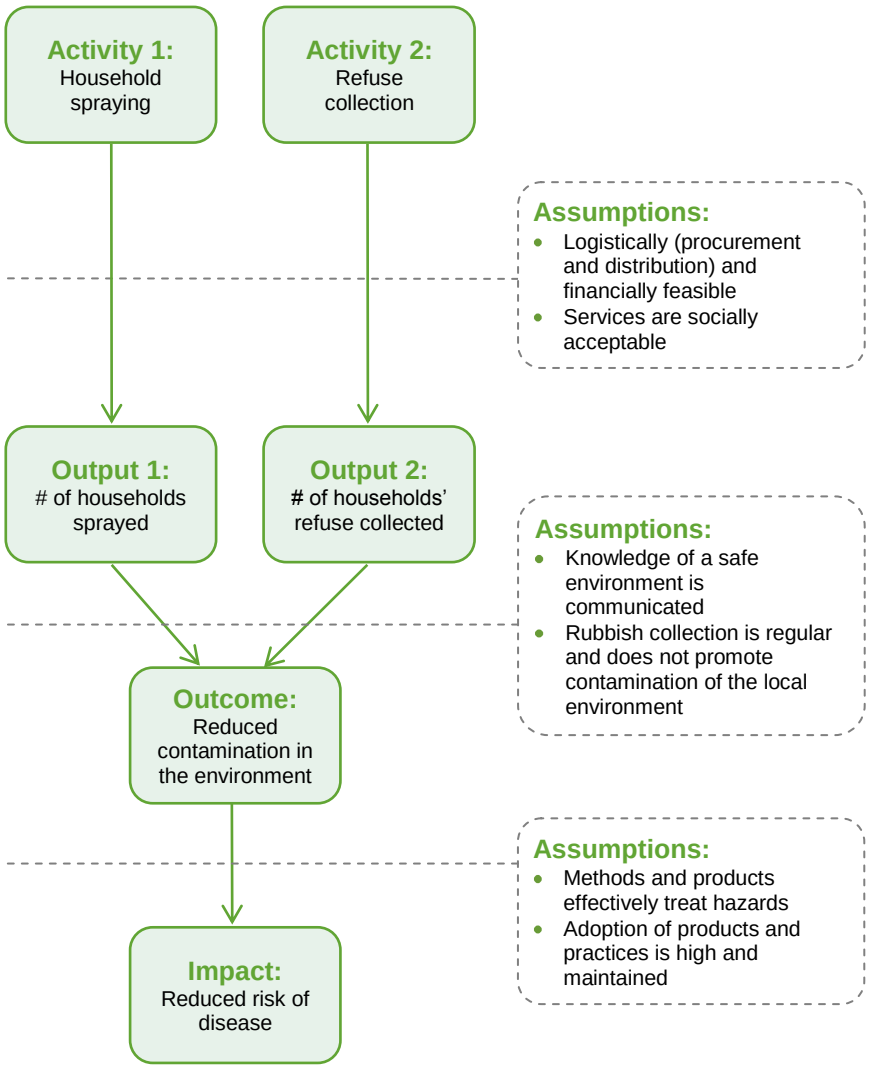
Figure 12: Distribution of soap and/or hygiene kit program theory



6) Environmental hygiene interventions

Environmental hygiene intervention program theory is based on the assumption that living in a clean environment reduces disease risk. Some of the basic assumptions are founded on good sanitation and personal hygiene practices, like no open defecation and handwashing at critical times. Education of households on routes of contamination relies on behavior change and households wanting to adopt new practices. Cleaning materials, i.e. chlorine solution, may have limited effectiveness if used on dirt floors or non-durable surfaces.

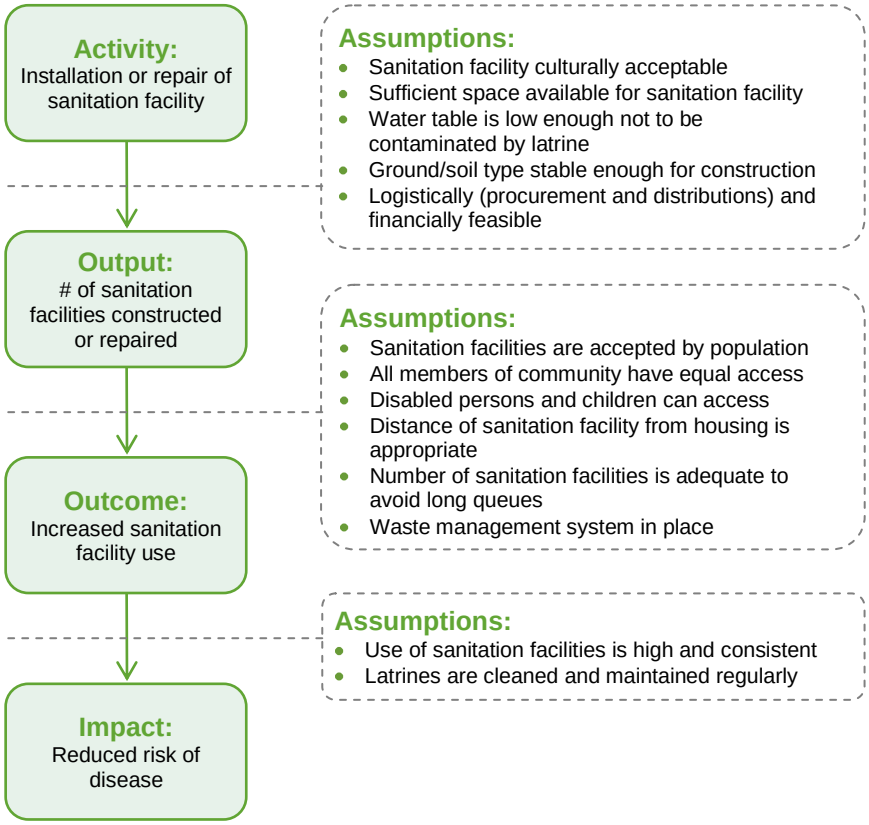
Figure 13: Environmental hygiene program theory



7) Installation of temporary or permanent sanitation facilities

The installation of sanitation facilities (i.e. latrines) program theory, assumes that there is adequate and available space to install sanitation facilities close to living quarters but are not a potential contamination hazard. The soil type and depth of the water table must also be considered as critical assumptions. Further behavior change activities, like hand-washing and no open defecation, are critical assumptions needed to make an impact.

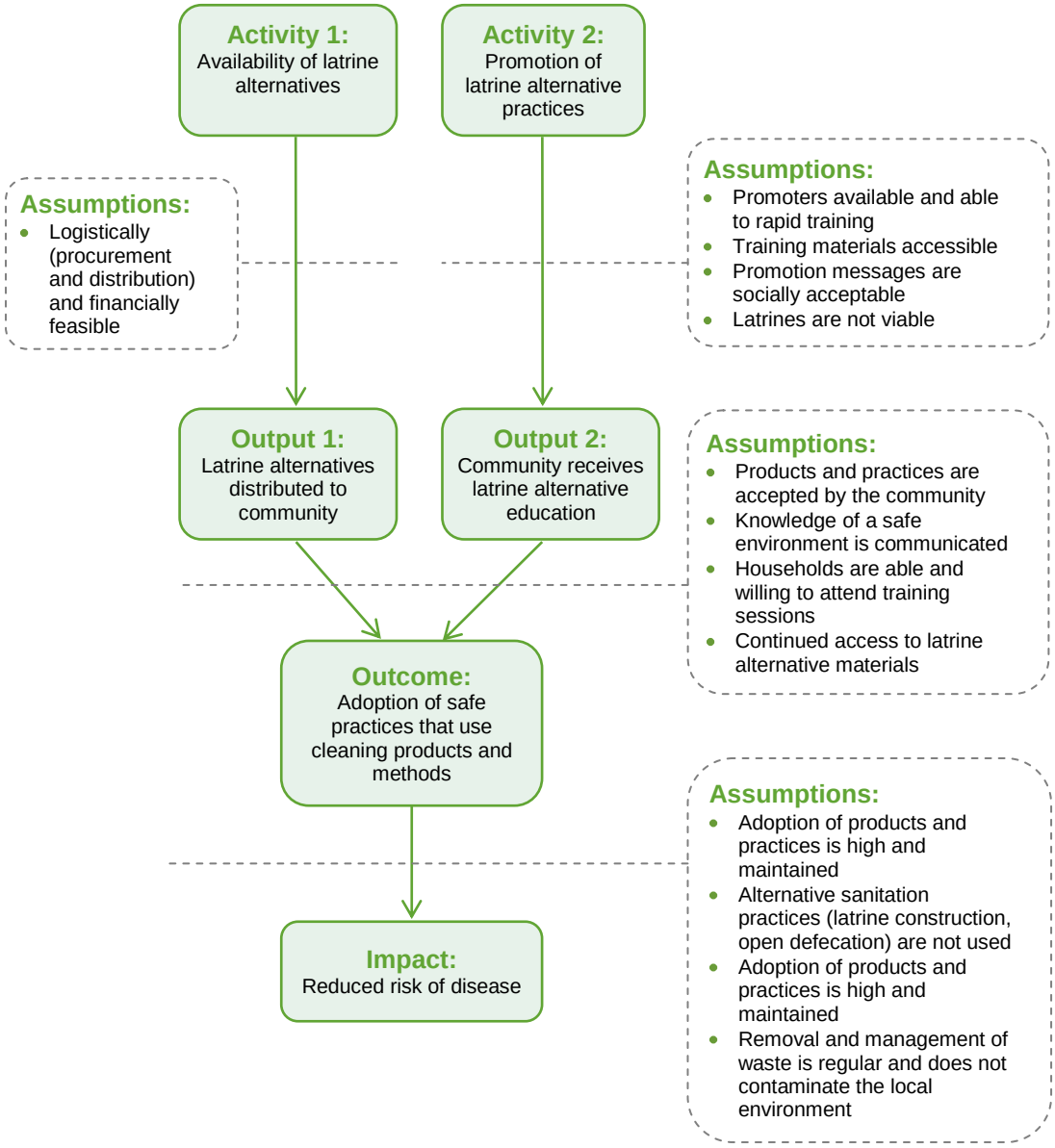
Figure 14: Installation of sanitation facility program theory



8) Distribution and management of latrine alternatives

Latrine alternatives are used in situations where latrines are not a viable option or will take too long to construct. The critical assumptions are that people will use the alternatives (with suitable training), but that there is a collection system that removes the waste from the household and is disposed in a safe place. The relatively new technology may limit the access in remote locations or willingness to move away from traditional sanitation actors.

Figure 15: Latrine alternative program theory



1.5 CONTEXT, HETEROGENEITY, AND MIXED METHODS

To assess a wide range of interventions on a global scale, “Understanding context is crucial for understanding [intervention] impact” [30]. The differences in outbreak contexts are foundational to this review. Previous knowledge of the intervention, existing behaviors, or type of outbreak are just some of the contextual factors that can carry a large influence. For example, similar chlorine Dispensers interventions carried out in four different emergency contexts (cholera in Sierra Leone, food security in Senegal, cholera in the Democratic Republic of Congo, and cholera in Haiti) resulted in a wide range of effective use (0-81%) [31]. There is no ‘silver bullet’ of WASH programs that is applicable in all situations [32]. A dedicated portion of the data extraction will therefore be focused on contextual factors that affect the WASH interventions like: disease type, displaced population, geographic region, urban/rural setting, training components, concurrent emergency, complimentary interventions, impact, and other characteristics.

The heterogeneity of the interventions, contexts, and outbreak-affected populations are expected to be high. The type and quality of research will also vary considerably. Data collection from the studies will be extensive in an effort to maximize the potential for comparisons during data analysis.

This review will greatly benefit from the use of mixed methods. As described above, the analysis will include a variety of sources, from peer-reviewed journals to grey literature. These will include experimental, quasi-experimental (i.e. case control), and non-experimental methodologies utilizing counter-factual and factual evidence. Counterfactual studies are those that establish impact by comparing two or more groups found in experimental or quasi-experimental evaluation designs. These study designs help to minimize bias and can often better establish intervention impact by controlling for various factors [30]. Factual analysis compliments the impact analysis of comparison studies by following the causal chain logic described above. Investigating key assumptions along the chain establish the success or failure of an intervention. Qualitative research will incorporate interviews and focus groups, highlighting the opinions and feelings toward interventions that are difficult to estimate in quantitative research. Investigating cost-effectiveness also expands the assessment by adding another lens to view WASH programming during disease outbreaks[33], [34].

Qualitative research and qualitative information will both be collected for this review.

Qualitative research is a research design that often involves interviews, focus group discussions, or simple observation. The information gathered is typically coded into themes and summarized as general thoughts and opinions of the persons involved.

Qualitative information, we define as ***contextual information***, will also be collected. Contextual information is the descriptive details that are important for this review; such as: country, type of disaster, implementation agency, and so on. Contextual information will be collected from both quantitative and qualitative research design of studies that meet the inclusion criteria.

2 OBJECTIVE OF THE REVIEW

This systematic review has a singular overarching objective in assessing the impact of emergency hygiene interventions. The primary research question will be answered through four secondary objectives that further evaluate: a) use of service and disease reduction, b) positive intervention characteristics; c) cost-effectiveness, and d) non-health related factors of emergency WASH interventions in disease outbreaks.

We consider 'context data' information which could be descriptive information from the studies, quantitative, or qualitative data not necessarily related to the research objectives but will enable a clearer assessment of homogeneity for analysis (e.g. country, disease type, setting). Contextual factors are not in the inclusion criteria, as they will be collected only after the selection of the studies.

Primary research question:

What are the **outcomes and impacts** of WASH interventions during disease outbreaks in humanitarian emergencies in LMIC?

Secondary research questions:

a) How do the use of WASH interventions reduce disease outbreaks? (quantitative analysis)

b) What are the program design and implementation characteristics that are associated with more effective programs? (qualitative analysis)

c) What is the cost-effectiveness of WASH interventions in emergency outbreak situations? (quantitative analysis)

d) What are the non-health related outcomes (i.e. psycho-social, quality of life, behavior change) affects from WASH interventions in emergency WASH interventions? (qualitative analysis)

To meet these objectives, a systematic process is described to identify and select studies in Section 3. Section 4 describes the methods of data extraction and synthesis that will be used to establish impact of emergency hygiene programs.

3 SELECTION OF MANUSCRIPTS

Manuscripts in this review meet specifications defined by the following PICOS protocol for inclusion criteria. Search methods for peer-reviewed and grey literature are described in section 3.2 and the selection process is explained in section 3.3.

3.1 CRITERIA FOR INCLUDING STUDIES IN THE REVIEW [PICOS]

Defining a priori the Populations, Interventions, Comparisons, Outcomes, and Study Types (PICOS) increases the transparency as to how and why studies were selected. The PICOS variables for this review are described in detail below.

3.1.1 Populations

Populations considered in this review are outbreak-affected that are also in a LMIC defined by the World Bank at the time the outbreak occurred. High income countries (HIC) are not within the scope of this review because the resources available for the response vary significantly from LMIC; in the case that there are valuable lessons to be learned from a HIC outbreak, we will make note in the narrative, but not include it in the impact analysis. All age, gender, and socio-economic demographics will be considered. Diseases can be endemic in some populations, but not in others, thus a simple number of cases cannot be used as a definition. For this analysis, we define an 'outbreak' in accordance with the WHO as either:

- The occurrence of disease in excess of the normal baseline (2 times the baseline) or a sudden spike in cases (2 times the incidence of new cases); or
- A single case of a communicable disease long absent from a population, or caused by an agent (e.g. bacterium or virus) not previously recognized in that community or area; or
- Emergence of a previously unknown disease [35]; or
- Any case of particular diseases of interest (cholera, Ebola and Hepatitis E).

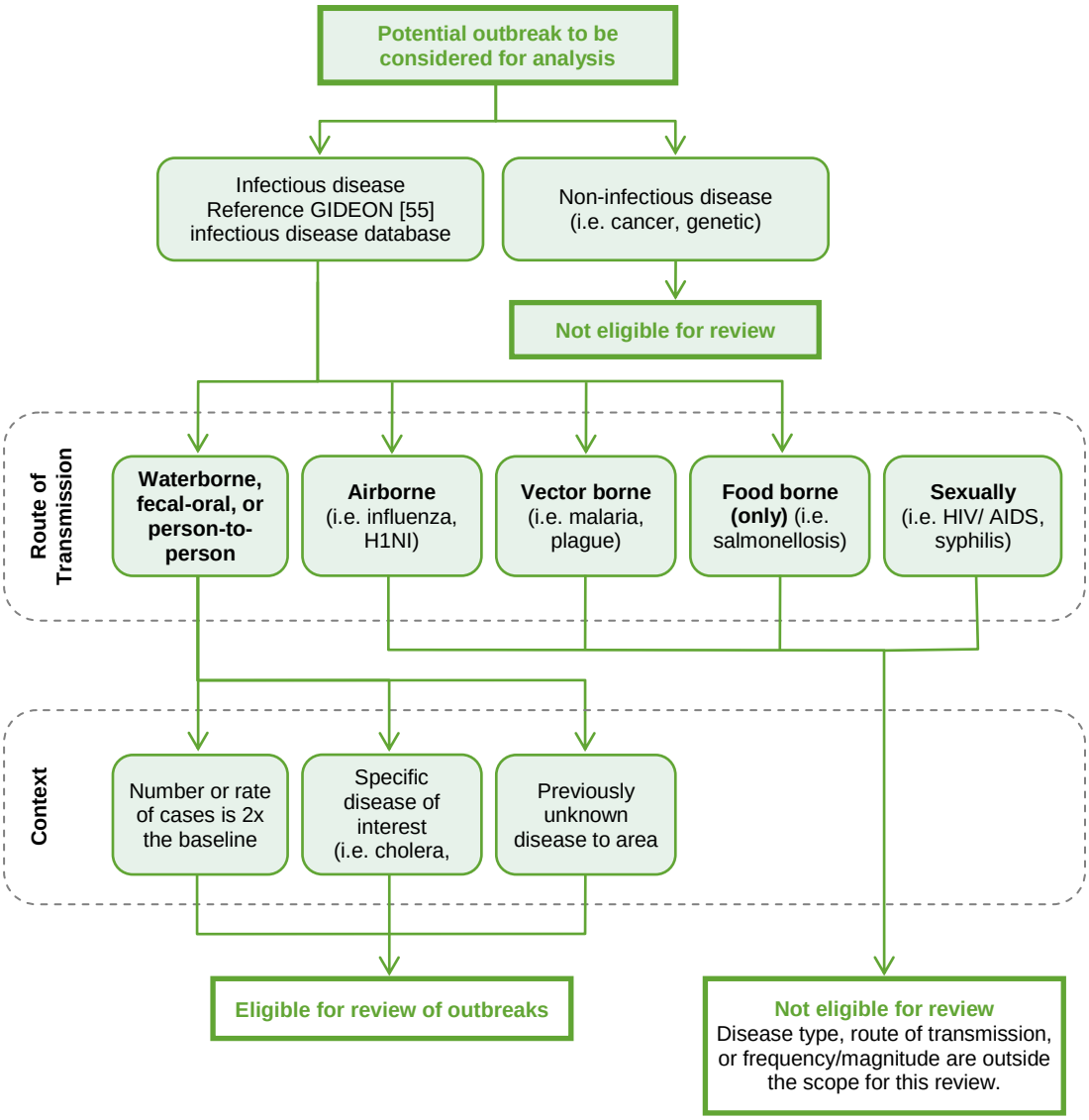
The WHO maintains a list of known outbreaks by disease type, year, and country dating back to 1996 [36]. The WHO outbreak list will be foundational in identifying outbreaks included in this review, but in situations that could be unreported or contexts are difficult to identify, a flow chart was developed to help differentiate contexts eligible for review (Figure 16). The flow chart is intended to assist in identifying an outbreak, but expert opinion and discussion of the research team will also be used. The infectious disease database, Global Infectious Disease and Epidemiology Online Network (GIDEON) [37], is the baseline information for the decision tree. Additional criteria for selection of eligible outbreaks relate to communicable diseases most relevant to WASH interventions. Outbreaks of interest are limited to common waterborne and fecal-oral diseases, as well as Ebola. Ebola is not a waterborne or fecal-oral disease, but important for review following the large-scale outbreak in western Africa. Not all waterborne or fecal-oral diseases could be assessed, thus, this review is limited to the seven diseases listed in Table 1. These diseases were selected because they are relevant to current WASH practitioners or are common diseases where WASH interventions may break multiple transmission routes. It is possible that WASH interventions could assist in prevention or control of other transmission routes or vectors; however, they are not eligible for review. Specific transmission routes not eligible for review include: vector borne (e.g. malaria, Dengue); airborne (e.g. influenza, H1N1); foodborne (e.g. food related salmonella); and blood/sexually transmitted (e.g. Hepatitis C, HIV) (Figure 16).

Table 1: Included Diseases

Communicable Disease	Transmission	Possible WASH Management
Bacillary dysentery (shigellosis)	Water/food and person to person	Safe water, sanitation, and hygiene
Diarrhea	Water/food, fecal-oral	Safe water, sanitation, and hygiene
Cholera	Water/food, fecal-oral, person-to-person	Safe water, sanitation, and hygiene
Hepatitis A	Fecal-oral	Sanitation and hygiene
Hepatitis E	Fecal-oral	Sanitation and hygiene
Typhoid Fever	Fecal-oral	Safe water, sanitation, hygiene
Ebola	Person-to-person	Precautionary personal hygiene measures, local (HH or clinic) environmental control

Adopted from Connolly (2005) [38].

Figure 16: Outbreak decision tree



3.1.2 Interventions

Inclusion for interventions fall into one of the eight interventions of interest: 1) increasing water access; 2) source-based water treatment options; 3) distribution of household water treatment technologies; 4) hygiene promotion; 5) distribution of soap and/or hygiene kits; 6) environmental hygiene interventions; 7) installation of temporary or permanent latrines; and 8) distribution and management of latrine alternatives.

The WASH interventions for inclusion must also directly target the prevention (i.e. bucket chlorination or HWT during a cholera outbreak) or control of disease transmission (i.e. chlorination of surfaces in an Ebola treatment unit). Also, interventions related to Ebola in West Africa, Hepatitis E in refugee camps, and cholera in new regions are of particular interest for review because of the immediate relevance in outbreak response; however, other infectious diseases are eligible for review (e.g. typhoid, dysentery) as described above.

3.1.3 Comparisons

As many relevant comparisons will be made to the best of ability of the data set. The eight interventions' impact will be compared with each other depending on intervention and control groups. Cost-effectiveness comparisons will also be incorporated into the analysis.

Factual evidence will be used to stratify the studies by the three primary manuscript types (peer-review, agency papers, and grey literature) as well as other WASH factors like: disease type, displaced population, geographic region, urban/rural setting, training components, concurrent emergencies, complimentary interventions, impact, and other characteristics.

3.1.4 Outcomes

A study would be included in the review if it reported on at least one intermediate outcome or final impacts that corresponds to the research questions in Section 2. Note: the program design characteristics are not inclusion criteria, but will come from contextual information collected from studies that also have at least one of the following outcomes or final impacts:

Intermediate Outcomes:

a) **Use of service:** Use of services is a general term that includes three specific definitions for: self-reported use, confirmed use, and effective use.
   a. Self-reported use is when a beneficiary reports the use of a product or event without additional verification. For example, self-reported use could be the recall of diarrhea episodes or daily use of a household treatment product. Self-reported use is often heavily biased.
   b. Confirmed use is when the evaluation tests, observes, or confirms in some way a product or service is used. For instance, testing free chlorine residual (FCR) in household drinking water 'confirms' the use of a water treatment method regardless of what the beneficiary reports.
   c. Effective use is the percentage of households improving their environmental hygiene quality from contaminated to uncontaminated by using a particular intervention; it combines both methods of confirmed use (through FCR or microbiological testing) as well as self-reported the use of the intervention.

b) **Economic analysis:** The outcomes collected for the economic analysis will include quantitative research and may include:
   a. Cost-benefit analysis;
   b. Cost-utility analysis;
   c. Cost per beneficiary; or
   d. Cost per Disability Adjusted Life-Year (DALY) averted.

Final Impacts:

c) **Disease reduction:** Morbidity and mortality reductions are the ultimate impact of the interventions. Assessing both the intermediate and final outcomes of the interventions allows the research team to evaluate the critical gap on the casual chain between outcome and impact. Final outcome measures are likely limited to quantitative research with several potential measures that are often expressed as a comparison over time or with another group in the form of an odds ratio (OR) or risk ratio (RR). Prevalence is expressed as a percent (%) of the population with a particular disease, while incidence is a rate of new cases over a specified time period.

   a. Morbidity rates (OR, RR, or case rates);

   b. Mortality rates (OR, RR, or case rates);

   c. Prevalence (%); or

   d. Incidence rates (cases/time).

d) **Non-health outcomes:** The non-health related outcomes could be from qualitative or quantitative research. The subjectivity of thoughts or feelings through focus groups or household surveys may be assessed but difficult to verify or clearly express their true meaning. For instance, questions like, “Do you like the taste of your drinking water after using a certain treatment method?” or “Why do you wash your hands?” could be quantified through a percentage of households in a survey, but primarily serve as qualitative research valuable to understanding how or why some interventions could be better suited in some contexts over others.

   a. Use of service (sustained difference in action by the population due to promotion, product input or context);

   b. Quality of life and Psycho-social affects (i.e. populations felt safer, more time for other things, less discrimination);

   c. User or agency preference of different interventions.

3.1.5 Study types

Due to the policy relevant research objectives and anticipated small amount of experimental evidence, all methodological designs are eligible for review (experimental, quasi-experimental, non-experimental, mixed-methods, and qualitative). Economic or cost analysis data will be included as dedicated studies or if it is specified as a component of broader research. Economic analysis could be cost-benefit analysis, cost-efficiency, or cost per DALY averted.

Initial scoping and previous research into WASH interventions in response to outbreaks yielded few experimentally designed evaluations from peer-reviewed journals. The majority of information is from quasi-experimental and non-experimental studies or grey literature. Some outbreaks have good WASH quasi-experimental information (such as case-control studies to identify risk and protective factors for cholera), however, other interventions, like handwashing, have more qualitative and non-experimental evidence. In order to fully capture the policy-relevant information for all data sets, the primary sources of data for this review will therefore include: the little existing experimental data supplemented by quasi-experimental and non-experimental manuscripts, agency documents from the UN or government body, and grey literature from NGOs.

We consider *agency reports* as an internally reviewed publication intended for an international audience. For example, agency reports could be a monthly situation report from the WHO in Senegal, or a global analysis from the World Bank. We consider *grey literature* as reports from NGOs that could be but is not typically expected to be made available on high-access external websites. Grey literature reports, for example, could be a project-specific impact analysis intended for a narrow audience, i.e. donor report. Within agency or grey literature, there will be a large variation in the scale of studies (global analysis to specific village impact) which also reflects the heterogeneity in study designs and quality of methodology.

In lieu of the breadth of grey literature, we will specifically exclude: personal blogs, diaries, newspapers articles, magazine articles, and legal proceedings/court documents. Books and dissertations will not be specifically searched but may be included in the review. Also, systematic reviews that meet the inclusion criteria will not be included, but references of systematic reviews will be collected for independent review.

Climate change may influence more frequent and severe weather, but the emergency response intervention remains focused the immediate flood, drought, or other disaster; thus climate change is outside the intended scope of review. We will record if studies identify climate change interventions in the context data collection, but it will not be a condition to include a study.

In health research, case-control study design is common, witnessed in our scoping assessment with many cholera studies. It is expected that there will be sufficient number of case-control studies to give confidence highlighting casual-chain assumptions. Another comparison method yielding counterfactual data will be with water quality testing, as some studies collect *E. coli* data of treated and untreated water, before and after an intervention or in household untreated and treated water pairs.

3.2 SEARCH METHODS FOR IDENTIFICATION OF STUDIES

A comprehensive search strategy will identify published and electronic literature. Each intervention will have a unique search strategy. Sources will be searched using keywords appropriate to each intervention studied. For example, a keyword combination for household water treatment could be represented as: (or 'cholera' or 'outbreak') and ('household water treatment' or 'point of use' or 'point-of-use' or 'water treatment'). A complete [Comment: is this list really 'complete'? Although the terms look appropriate and wide ranging, isn't it likely that the list of specific interventions will expand as the searching gets under way and additional propriety or local names are found for some of the interventions?] list of keywords is included in **APPENDIX C: Keywords**. Keywords will be searched in ten electronic databases, including:

- Cochrane Library
- Google Scholar
- IDEAS
- LILACs
- Ovid Medline (Pubmed)
- Web of Science
- Academic Search Premier (French)
- ARTFL-FRANTEXT (French)
- ArticleFirst (French)

We have already consulted, and will continue to work with, Karen Vagts, a Tufts University librarian and information retrieval specialist, to finalize the search strings for the electronic databases. Additionally, the journals: Journal of Water and Health; Journal of Water, Sanitation, and Hygiene for Development; Disasters; Disaster Medicine and Public Health Preparedness; Prehospital and Disaster Medicine; and Waterlines will be manually searched for relevant manuscripts. For studies with a specified document date (e.g. date of publication), dates for inclusion will be 1995-2015, regardless of when the research took occurred. For example, a study carried out from 1993-1994 but only published in 1995 would be eligible for review. Searches will be conducted in the English, Spanish, and French; however, manuscripts in any language are eligible for review. Native speakers will be asked to volunteer their assistance in evaluating the eligible manuscripts not in English, Spanish, or French.

The identified limited number of quality peer-reviewed manuscripts increases the importance of unpublished grey literature. Grey data repositories, opengrey.org and greylit.org, will be searched in a manner similar to the peer-reviewed databases. A wide array of agencies will be approached through direct email solicitation and agency website searches (APPENDIX D: List of Websites and Organizations for Electronic Searches), representative examples include:

- **UN Agencies / International Bodies:** Unicef, WHO, UNHCR, OCHA, ICRC, IOM
- **Government agencies/Donors:** CDC, Office of Foreign Disaster Assistance (OFDA), EU Humanitarian Aid and Civil Protection (ECHO), Department for International Development (DFID), Humanitarian Innovation Fund (HIF)
- **Development Banks:** World Bank, Asian Development Bank, African Development Bank, Inter-American Development Bank
- **WASH Networks:** the WASH Cluster email list, the WASHPlus email list, Active Learning Network for Accountability and Performance (ALNAP), Delft and Water Engineering and Development Centre (WEDC) university programs
- **Private foundations:** The Bill and Melinda Gates Foundation, The Clinton Foundation
- **NGOs:** Action Against Hunger, Medecins Sans Frontieres, Oxfam, International Rescue Committee, Save the Children

Websites often have less search capabilities than electronic journals. To address this, we will work with the information retrieval specialist to customize the searches specifically for websites. Reference snowballing will also be completed, particularly in reaching out directly to authors of reports and authors in the reference list who might have additional unpublished information. Systematic reviews will not be included in this research; however, references from systematic reviews that meet initial screening criteria will be used to collected for individual inclusion. References from manuscripts that meet the full inclusion criteria, described in Section 3.3, will also be evaluated for inclusion.

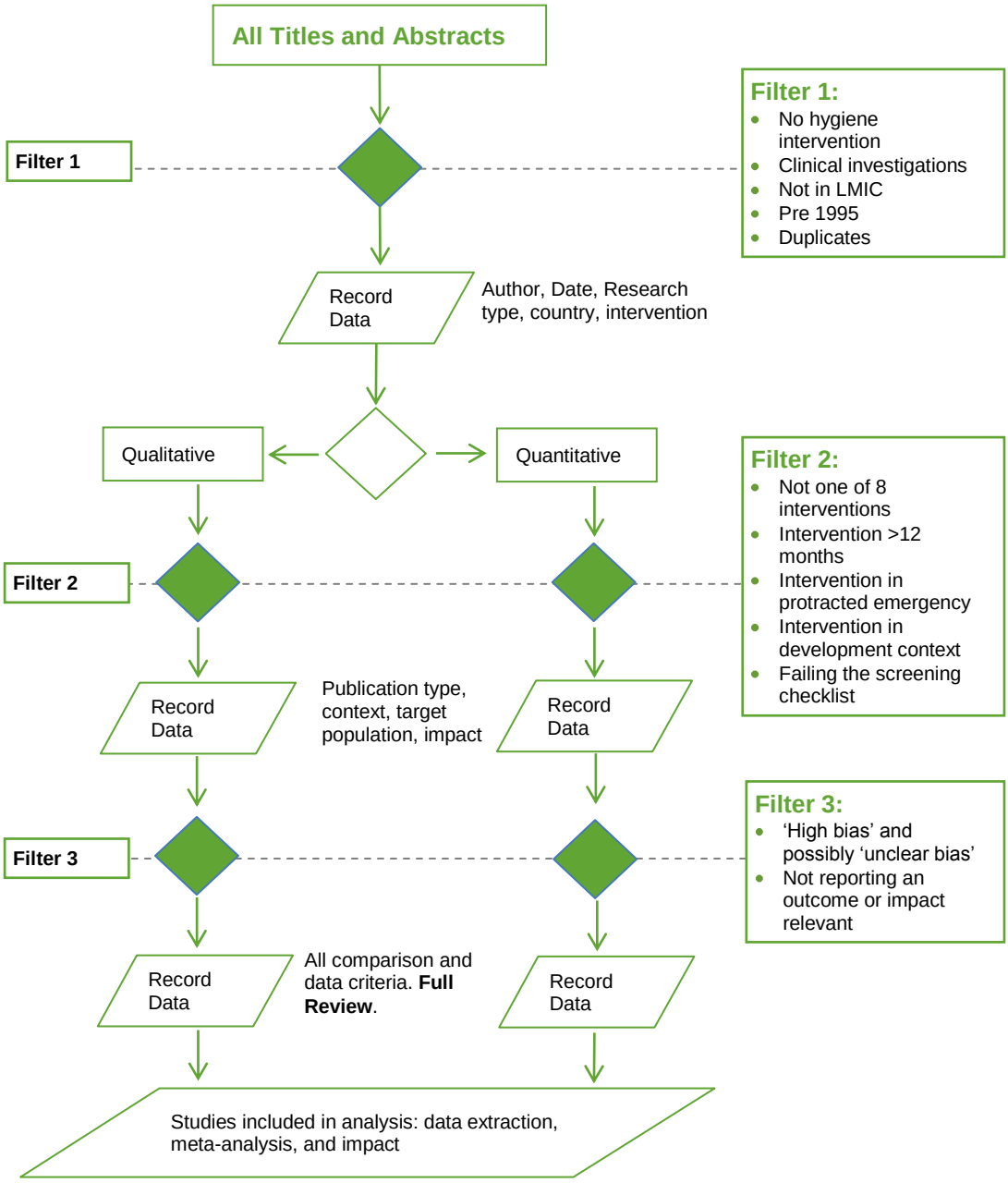
3.3 SELECTION OF STUDIES

The selection of studies will adhere to the principle standards of the Cochrane Intervention Reviews [39]. All gathered titles/abstracts will be numbered in sequence for identification to begin the three stage selection of studies. To achieve independent double screening after the initial title/abstract filter, two team members will review the manuscripts for stage 2 and 3 of the selection process. For stage 2, a research assistant and Mr. Yates will double screen the studies. On the final filter, one of the three hygiene experts will be the primary reviewer, with Mr. Yates acting as a secondary reviewer. A summary of the selection process is described in Figure 17 with more detailed description of each stage below.

Filter 1: Filter 1 will exclude the following studies:

1) No water, sanitation, hygiene, environmental intervention (very liberal definition).
2) Clinical or hospital diagnoses will be eliminated because there is no intervention and non-communicable diseases will be eliminated because it is outside the scope.
3) Not implemented in a LMIC as defined by the World Bank. This will exclude studies in the United States of America, Canada, Western Europe and other developed nations.
4) Studies published before 1995.
5) Duplicates.

Figure 17: Summary flow diagram for study selection process



Filter 2: The downselected titles/abstracts will be coded only by type of most relevant hygiene intervention then reviewed by a research assistant and Mr. Yates for more stringent criteria. Exclusions for filter two result if *any* of the following are true:

1) Study not evaluating one of the eight types of hygiene interventions;
2) Interventions of more than 12 months.
3) Interventions in a protracted or chronic emergency.
4) Interventions in a development context.
5) Studies that fail the checklists in Appendix E. Short checklists for various quantitative studies, as well as, qualitative and economic studies will help identify weak studies without a full review. Each of these criteria will be coded in the master Excel spreadsheet.

Abstracts will be included in the full analysis if **one** or **both** reviewers support inclusion. Full studies will be downloaded then reviewed by Mr. Yates and one of the hygiene experts (Table 1).

Table 2: List of Reviewers for Each Hygiene Intervention (Filter 3)

Hygiene Intervention	First Reviewer	Second Reviewer
Well Rehabilitation	Dr. Lantagne	Mr. Yates
Source Based Water Treatment	Dr. Lantagne	
Household Water Treatment	Dr. Lantagne	
Hygiene Promotion	Ms. Vujcic	
Hygiene Kit Distribution	Ms. Vujcic	
Environmental Hygiene	Ms. Vujcic	
Latrine Installation	Dr. Joseph	
Latrine Alternatives	Dr. Joseph	

Filter 3: The two reviewers will evaluate the studies to independently assess the reported outcome, impact, or assessment that is relevant to a hygiene intervention OR qualitative information OR economic analysis.

During this process, the research team will assess potential for additional confounding factors, adherence to the scope of review, inconsistent outcomes or impact, unjustified conclusions and discuss any potential concerns with each other. Both reviewers must approve study for final inclusion. Any discrepancy will be determined by a third reviewer.

We do not expect an overwhelming amount of relevant studies that would be included in the review; however, given that possibility, we will remove manuscripts with the highest risk of bias score, Annex F.

If the revised number of relevant studies eligible for inclusion remains greater than 200, we will discuss possible options with 3ie and our advisory committee.

4 DATA EXTRACTION AND PROCESSING

Once the manuscripts have been down selected, data will be gathered for comparison and data extraction. A full list of criteria collected is listed in Appendix A. The comprehensive list of criteria will establish the underpinnings for comparisons and appreciation of heterogeneity of the studies.

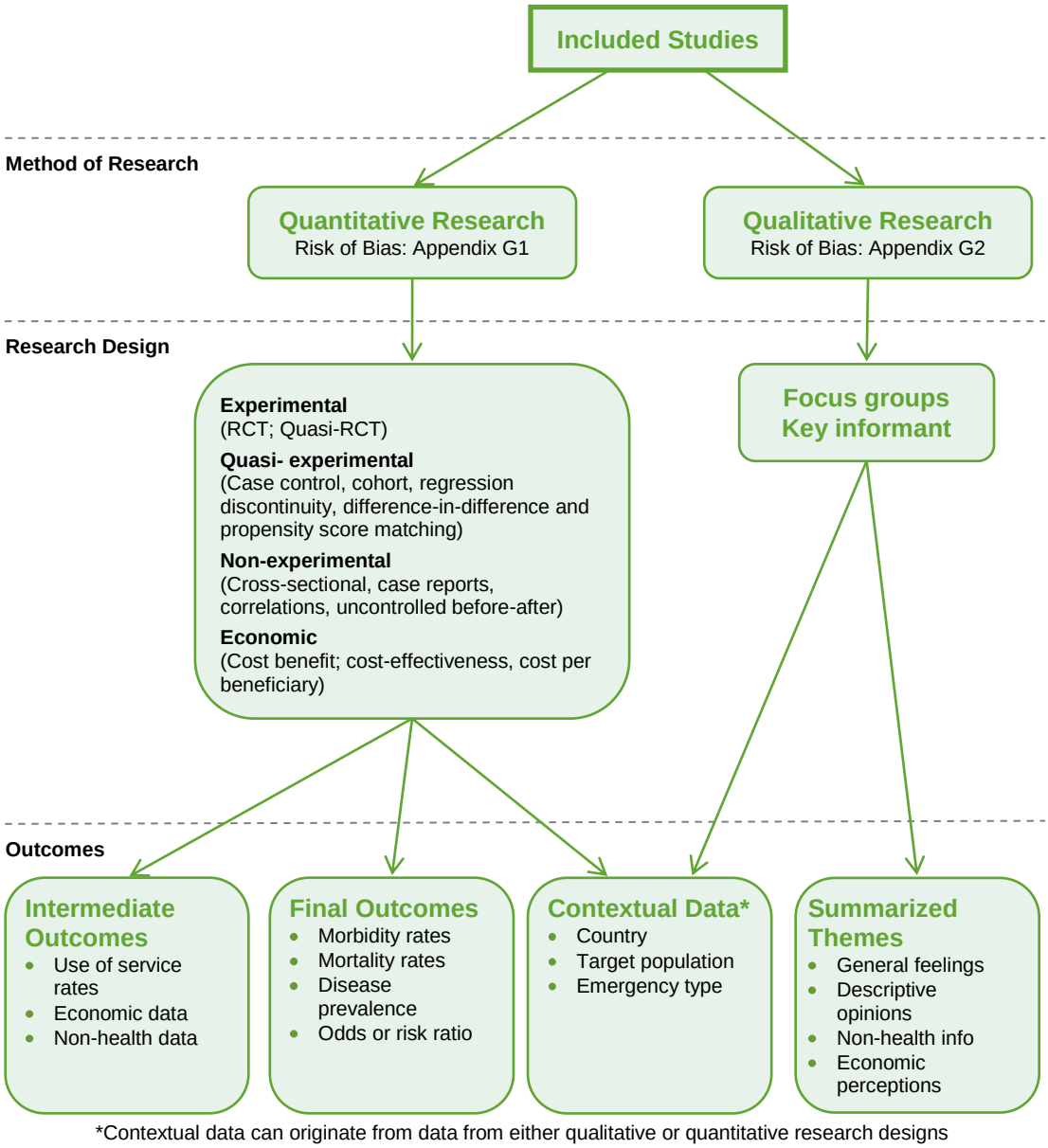
4.1 CODING

Studies included in the review will be coded by research assistants and the review team. The coding will be completed by a team of two to three people. Initially, the research assistants and Mr. Yates will review and code at least 10 studies as a group to establish consistency. Then the research assistants and Mr. Yates will code the remaining studies individually. Outcome measures will be double screened for accuracy by a member of the review team according to their expertise.

Information recorded from each manuscript is based on the Waddington et al. (2012) protocol and will describe: author and publication details, type of intervention, context of the intervention, study design, study quality, effect estimation, intermediate outcomes, qualitative information, economic outcomes, and final outcomes. Detailed criteria from all included studies (quantitative, qualitative, or economic) will be extracted into a master list in Microsoft Excel (2010).

From the initial screening, studies have been sorted into quantitative or qualitative research. Separating the studies by research method allows the data collection to address the differences in the types of research. Figure 18 is a descriptive flow chart of the types of studies expected in this review, with the different outcomes from the various study designs.

Figure 18: Source of Data Retrieval Flow Diagram



*Contextual data can originate from data from either qualitative or quantitative research designs

4.2 QUALITY APPRAISAL

The risk of bias tools are also separated by research design: quantitative and qualitative. Each tool summarizes a study into 'high risk,' low risk,' or 'unclear.'

4.2.1 Quantitative appraisal

To determine the risk of bias in quantitative studies (experimental, quasi-experimental, and non-experimental), an assessment tool was developed, based on the Cochrane Handbook Risk of Bias Tool while also drawing heavily on the structuring and description by Baird et al (2013). We will assess the risk of bias through five categories: 1) selection and confounding; 2) spillover and contamination; 3) incomplete outcome; 4) selective reporting; and 5) other bias. Similarly described by Baird et al. [40]:

- **Selection and confounding:** addresses the issue of program design. Allocations, selection of beneficiaries, targeting, and matching concerns are represented in this category.
- **Spillover and contamination:** addresses the issue of spillovers from the treatment to the control group. Not controlling for outside factors or additional interventions in the area also have spillover effects.
- **Incomplete outcome:** addresses the issue of whether analysis of all relevant outcomes was reported or whether there appears to be selection in reporting. Loss to follow-up or missing data can reduce the power of the research design as well as potentially introduce bias with unequal loss of sample between groups.
- **Selective reporting:** authors utilize a credible analysis method and report on all intended outcomes. Some research is funded by manufacturers of products, which can lead to selective reporting of only favorable outcomes.
- **Other risks of bias:** this category is to any number of other risks of bias present in the report. Self-reported data is of particular concern for our analysis. Also, retrospective baseline data, data using inappropriate methods, changing follow-up methods or procedures are examples of other potential biases. This is the most subjective of the five categories.

Each study will be scored across the five categories as 'Low Risk,' 'High Risk,' or 'Unclear.' The overall determination for the risk of bias for that study is assessed with the table below, summarizing the five categories into a single quality assessment for each qualitative study.

Table 3: Risk of Bias Summary

Risk of Bias	'Low Risk' Assessed in Categories
Low Risk	4-5 'Low Risk' Scores
Medium Risk	3 'Low Risk' Scores
High Risk	1-2 'Low Risk' Scores

4.2.2 Qualitative appraisal

The qualitative assessment has been adapted from Spencer et al. 2003 "Quality in Qualitative Evaluation: A Framework for assessing research evidence" [41], [42]. The quality assessment is evaluated on four appraisal questions. There is no clearly objective rule for determining bias among qualitative studies. The guiding questions will be used by the research team to help establish core research questions that should be evaluated; however, professional judgment is necessary to make the assessment. Qualitative experts are on the research team and advisory board to ensure rigorous standards, consistency, and transparency.

Each study will be scored across the four appraisal questions categories: 1) design; 2) bias; 3) data collection; and clarity of finding as 'Low Risk,' 'High Risk,' or 'Unclear.' The overall determination for the risk of bias for that study is assessed with the table below.

- **Design:** The overall design of the research is considered, especially the targeting of the research population.
- **Bias:** How representative is the research population compared and are there obvious biases that affect the findings?
- **Data Collection:** How was the data collected, recorded (audio, video, transcribed)? Who collected the information?
- **Clarity of findings:** Do the conclusions match what could be achieved from the study design? Is there an inherent logic to the conclusions?

Table 4: Risk of Bias Summary

Risk of Bias	'Low Risk' Assessed in Categories
Low Risk	3 or more 'Low Risk' Score
Medium Risk	2 'Low Risk' Scores
High Risk	1 or less 'Low Risk' Score

4.2.3 Economic appraisal

Economic assessments can be the primary purpose of the study or a component of a larger study. In either case, the economic review tool is a framework to assess the validity of economic information. The economic assessment tool (Appendix G3) is to be used in addition to the quantitative or qualitative tools found in Appendix G1 or G2. It was adapted from the CASP Economic Checklist [43]. Examples of economic studies could be cost-benefit or cost-effectiveness analysis. Simple cost statements or budget analyses will be recorded as contextual information, unless some formal economic evaluation was carried out.

4.3 MEASURES OF TREATMENT EFFECT

Data will be collected from the selected studies including: sample size, 95% confidence intervals, and impact estimates. Where appropriate, the standardized mean difference will be determined for continuous variables, while odds ratios or risk ratios will be used for dichotomous variables. Impact will be described as a difference between groups (i.e. difference of means) or a ratio (i.e. risk ratio or odds ratio). As often reported in health studies, ratios less than 1.0 represent a protective effect, while ratios greater than 1.0 represent an increased risk. Data transformations will be conducted as necessary according to the most appropriate methodology.

Studies with that have effect sizes with more precision will have more influence for the overall effect in the meta-analysis by using the 1/(standard error2) for random effects variance. Additionally, small sample size correction and robust standard errors will be used when necessary as described by Baird et al [40].

4.4 MISSING DATA

Primary authors will be contacted for missing data. Where no additional data can be retrieved, the use of response ratios will be used as outlined by Waddington et al. (2012) and further described by Borenstein et al. (2009)[44], [45]. The response ratio measures the change between intervention and control groups by a simple proportion, similar to a risk ratio.

$$R = \frac{x^t}{x^c}$$

where R is the response ratio effect, X_t is the mean outcome of the intervention group, and X_c is the mean control group [45]. The response ratio described above may be used to compare different study designs with similar outputs. Waddington (2012) describes that due to the response ratio comparing effect only, difference-in-difference designs or propensity scoring designs can be compared side-by-side. Odds ratios may be converted to effect size in accordance with Chinn (2000) [46]. Studies without control groups or datasets where a response ratio cannot be used, baseline information will be used; if comparison is not possible, then results will be reported qualitatively.

For qualitative research, we will also request the authors to provide primary data transcripts of the key informant interviews, focus group discussions, or other data collected. All reasonable attempts to include missing data will be made; however, given the timeframe

allotted for analysis and reporting, this may not be possible. If missing data is thought to jeopardize the deliverables, the studies will be documented, but removed from analysis after discussion with the advisory board and HEP.

4.5 UNIT OF ANALYSIS ISSUES

Issues can arise when studies collect randomized information at an individual level (i.e. household) through geographic clusters (i.e. village). In situations where differences between the clusters are greater than differences within a cluster, the confidence intervals are incorrectly small (Waddington et al. 2012). This is a result from violating the assumption that comparisons within the cluster (village) are independent. Studies that do not fully control for this clustering effect have a unit analysis error that will be corrected. Standard error and confidence intervals will be adjusted with original data or an intra-cluster correlation coefficient of 0.02 will be used to make corrections.

4.6 HETEROGENEITY ASSESSMENT

The anticipation of heterogeneity is the catalyst for the comprehensive collection of context and study criteria. Careful consideration will be made to appreciate the heterogeneity and implications of results, with respect to statistical characteristics (sample size, power) and generalizability. We will ensure to note which included data in each of the stratifications comes from which sources. With qualitative or less comparable data, we will clearly express the limits of any external comparisons.

Heterogeneity will be assessed with up to three methods: Cochrane's Q, Tau^2 and I^2. Generally, more weight will be given to Tau^2 and $I^{2;}$ however, rationale for establishing or rejecting heterogeneous conclusions will be stated when tests contradict. Contextual factors from qualitative data will be included to understand the variation in results, as research is clear that intermediate outcomes vary significantly between contexts.

Example groupings are: time since the onset of the outbreak, training components, displacement of the population, outbreak occurring after an emergency or not, outbreak occurring in a new context or in a context where disease in known, urban/rural setting, geographic region, and complementary interventions.

4.7 SUBGROUP ANALYSIS

In interventions with sufficient data, sub-group analysis will be completed by stratifying the data into relevant groupings. Subgroup analysis will follow the PROGRESS-Plus criteria. These subgroups comprehensively differentiate subsets of the general population that are often vulnerable or discriminated against. A portion of the data collection variables are dedicated to PROGRESS-Plus categories; however, given the type of research carried out in emergencies, it is expected that only age and gender subgroups are expected for subgroup analysis. If additional subgroups become apparent, we will provide further analysis. We will clearly state which manuscripts are included in each stratification group.

4.8 METHOD OF SYNTHESIS

We will synthesize outcomes across programs, considering contextual factors, timing of interventions, and training provided to recipient population. Stata statistical software will be used for data analysis.

Meta-analysis techniques (e.g. weighted average, pooled effect, forest plots, and funnel plots) for outcome assessments will be pursued if sufficient experimental design studies meet study inclusion criteria. Forest plots will be most useful to display the range of effect sizes across

the findings [39], [47]. Difference in the timing of interventions could be a unique analysis regarding the time between the onset of a disaster and different interventions, with effect size presumably changing over time. We would also like to assess the length of time before a particular outcome or impact is achieved; however, this is not expected to be possible with most interventions of interest. Improvements in water quality will likely be one area where significant synthesis can occur. Before synthesis, we will critically evaluate the quality of water quality testing in each of the studies to determine if *E. coli* or thermotolerant coliform data can be included in the calculations. Case-control data, particularly from cholera outbreaks, is another likely source of data that can be statistically analyzed.

The response ratio described above may be used to compare different study designs with similar outputs. Waddington (2012) describes that due to the response ratio comparing outcome effects only, some quasi-experimental designs can be compared side-by-side. We will also highlight outcome effect consistency to determine expected impact and relevance. Consensus among the review team with oversight from the Advisory Board will determine a level of confidence in each intervention as low/moderate/high to help guide policy and future research.

Qualitative Synthesis

We will combine related qualitative research material into file sets, and re-code data (if necessary) using qualitative analysis program Atlas.ti. We will review the codes to develop themes that reflect the gaps in the causal chain and then develop qualitative result summaries based on the themes. Direct quotations will be used to highlight key results. Qualitative research will be used to evaluate the gaps in the casual chains through factual analysis.

Economic Synthesis

Cost-effectiveness will be assessed using the range of 1-3 times the per capita income for the country of intervention [48]. Studies that have economic or cost-effectiveness outcomes, we will use the CASP economic checklist to help synthesize data along with guidance from the WHO Manual for Economic Assessment of Drinking Water Interventions[43], [49]. Results will be standardized to common metrics, such as $/DALY averted or cost per user, and compared across interventions. Costs will be normalized and converted to 2015 USD. Simple costs per beneficiary metrics will be considered high risk, unless there are clear descriptions about what is included in the analysis.

Integrated synthesis

This comprehensive review makes use of qualitative, quantitative, and contextual factors. By assessing all three data sources, an integrated synthesis of the causal chain can be evaluated. We will combine and contrast data from all three data sources to have a more robust understanding of the emergency hygiene interventions. This evaluation will shed new light on how the humanitarian response community views the emergency hygiene causal chain, potentially influencing how future programming is implemented or guiding future work in the sector.

4.9 SENSITIVITY ANALYSIS

Sensitivity analysis will assess the risk of bias, study design type, treatment effect, and possible outliers. Hard cut offs for exclusion criteria are minimized, but if required, the researchers will discuss and agree upon an excepted level. Rational for inclusion or exclusion of a study will be recorded to facilitate the sensitivity analysis to ensure all appropriate studies are included in the analysis.

Examining the eight WASH interventions individually will help to narrow assumptions made in the causal chain. Case studies, as well as, including relevant grey literature and qualitative studies will also help to identify contextual factors of the interventions and potential implementation hurdles that break the assumed causal chain.

4.10 DEPENDENCY OF STUDIES

The unit of observation for this review is on the intervention level, thus we will construct one effect size for each intervention in each study according to the outcomes of interest. There will likely be the case where multiple studies report on similar interventions, but from different NGOs in a particular emergency, or similar interventions by one NGO but in different emergencies in a single study; in either case, both studies will be assessed. In the situation where an NGO report is followed by a white paper or journal article, we will include only one study with the lowest risk of bias. Similarly, when dissertations and journal articles overlap in content, only one study will be included. If the risk of bias is the same, then inclusion will be made on if it was or to the level of being peer-reviewed.

Where possible, sub-groups will be analyzed by outcomes. it is likely that we will *synthesize* and *summarize* the same data set several times, following the methodology described by Baird et al. 2013 [40].

Baird et al. describes *synthetic effects* from non-independent data; studies that use the same populations with several different interventions or outcomes. Synthesis is simply the average effect size, with the correlation coefficient assumed to equal 1.0, representing the variance of the mean. *Summary effects* are when studies are independent and subgroup effect size is often reported. A random effects model will be used to combine effect size for independent studies. Forest plots will be utilized for graphic representation of the summary data. Replication of research with the same population will be included and analyzed independently.

Where the studies are assessed as independent with sufficient information, subgroup analysis for meta-analysis will be carried out. When individual studies report on multiple outcomes, we will attempt to summarize one outcome from the study according to each of our outcomes of interest. Where multiple interventions are carried out simultaneously and assessed together, secondary analysis will assess the difference in effect size of individual interventions, indicating potential synergies.

4.11 SUMMARY OF FINDINGS

The quality of evidence assessment for this review will be summarized with the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach. GRADE is outlined in the Cochrane Handbook in chapter 12 as a way to evaluate summary findings with respect to effect size, research design, and bias. A summary table for each of the interventions and subgroups will be created with expected effects and confidence in the results. Additionally, forest plots will be used to display effect sizes graphically, and funnel plots will display potential for publication bias.

For the final report, interventions will be grouped or clustered to most appropriately display the data assessed. This may not necessarily be aligned with the eight interventions described above, but may be grouped to be most relevant for field practitioners. Groupings and the display of results will be made with suggestions from the advisory committee while keeping a mindset of policy relevance and usability for humanitarian actors.

5 ACKNOWLEDGEMENTS

We would like to thank the Humanitarian Evidence Program along with Oxfam Great Britain, Feinstein International Center at The Friedman School of Nutrition Science and Policy at Tufts University, and UK Aid for the opportunity and funding to conduct this research.

6 REFERENCES

[1] "Disease outbreaks," *The World Health Organization*, 2015. [Online]. Available: http://www.who.int/topics/disease_outbreaks/en/. [Accessed: 14-Sep-2015].

[2] K. F. Smith, M. Goldberg, S. Rosenthal, L. Carlson, J. Chen, C. Chen, and S. Ramachandran, "Global rise in human infectious disease outbreaks.," *J. R. Soc. Interface*, vol. 11, no. 101, p. 20140950, 2014.

[3] P. Greaney, S. Pfiffner, and D. Wilson, Eds., "Sphere Handbook: Humanitarian Charter and Minimum Standards in Disaster Response," The Sphere Project, Geneva, Switzerland, 2011.

[4] J. T. Watson, M. Gayer, and M. a Connolly, "Epidemics after natural disasters.," *Emerg. Infect. Dis.*, vol. 13, no. 1, pp. 1–5, Jan. 2007.

[5] "WHO / UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation," 2014. [Online]. Available: http://www.wssinfo.org/data-estimates/tables/.

[6] "UNOCHA 2014 Year in Review," 2014. [Online]. Available: www.unocha.org/2014_year_in_review/. [Accessed: 15-Jul-2015].

[7] M. J. Toole, "Mass Displacement. A Global Health Challenge.pdf," *Infect. Dis. Clin. North Am.*, vol. 9, no. 2, pp. 353–366, 1995.

[8] Goma Epidemiolgy Group, "Public Health Impact of Rwandan Refugee Crisis," *Lancet*, pp. 339–345, 1995.

[9] E. H. Teshale, C. M. Howard, S. P. Grytdal, T. R. Handzel, V. Barry, S. Kamili, J. Drobeniuc, S. Okware, R. Downing, J. W. Tappero, B. Bakamutumaho, C.-G. Teo, J. W. Ward, S. D. Holmberg, and D. J. Hu, "Hepatitis E Epidemic, Uganda," *Emerg. Infect. Dis.*, vol. 16, no. 1, pp. 126–129, Jan. 2010.

[10] A. Bastable and L. Russell, "Gap Analysis in Emergency Water, Sanitation and Hygiene Promotion," 2013.

[11] M. J. Toole, "Mass population displacement. A global public health challenge," *Infectious Disease Clinics of North America*, vol. 9, no. 2. pp. 353–366, 1995.

[12] M. J. Toole, "Vulnerability in emergency situations.," *Lancet*, vol. 348, no. 9031, p. 840, Sep. 1996.

[13] M. A. Connolly, M. Gayer, M. J. Ryan, P. Salama, P. Spiegel, and D. L. Heymann, "Communicable diseases in complex emergencies□: impact and challenges," *Lancet*, vol. 364, pp. 1974–1983, 2004.

[14] M. J. Toole and R. J. Waldman, "The public health aspects of complex emergencies and refugee situations.," *Annu. Rev. Public Health*, vol. 18, pp. 283–312, Jan. 1997.

[15] W. J. Moss, M. Ramakrishnan, D. Storms, A. Henderson Siegle, W. M. Weiss, I. Lejnev, and L. Muhe, "Child health in complex emergencies.," *Bull. World Health Organ.*, vol. 84, no. 1, pp. 58–64, Jan. 2006.

[16] H. Waddington and B. Snilstveit, "Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea," *J. Dev. Eff.*, vol. 1, no. 3, pp. 295–335, Sep. 2009.

[17] P. R. Hunter, "Household Water Treatment in Developing Countries: Comparing Different Intervention Types Using Meta-Regression," *Environ. Sci. Technol.*, vol. 43, no. 23, pp. 8991–8997, Dec. 2009.

[18] T. Clasen, I. Roberts, T. Rabie, W. Schmidt, and S. Cairncross, "Interventions to improve water quality for preventing diarrhoea.," *Cochrane database Syst. Rev.*, vol. 3, no. 3, p. CD004794, Jan. 2006.

[19] L. Fewtrell, R. B. Kaufmann, D. Kay, W. Enanoria, L. Haller, and J. M. Colford, "Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis," *Lancet Infect. Dis.*, vol. 5, no. 1, pp. 42–52, Jan. 2005.

[20] M. Clarke, C. Allen, F. Archer, D. Wong, A. Eriksson, and J. Puri, “What evidence is available and what is required, in humanitarian assistance? 3ie Scoping Paper 1,” New Dehli, 2014.

[21] J. Parkinson, “A Review of the Evidence Base for WASH interventions in Emergency Responses,” 2009.

[22] J. Darcy, H. Stobaugh, P. Walker, and D. Maxwell, “The use of evidence in humanitarian decision making,” *Learn. Pap. Tufts Univ. Somerville, USA*, 2013.

[23] J. Darcy, H. Stobaugh, P. Walker, and D. Maxwell, “The Use of Evidence in Humanitarian Decision Making ACAPS Operational Learning Paper,” pp. 1–39.

[24] S.-L. Loo, A. G. Fane, W. B. Krantz, and T.-T. Lim, “Emergency water supply: a review of potential technologies and selection criteria.,” *Water Res.*, vol. 46, no. 10, pp. 3125–51, Jun. 2012.

[25] A. Steele and B. Clarke, “Problems of treatment process selection for relief agency water supplies in an emergency,” *J. Water Health*, vol. 6, no. 4, pp. 483–489, 2008.

[26] C. C. Dorea, “Comment on ‘Emergency water supply: A review of potential technologies and selection criteria.,’” *Water Res.*, vol. 46, no. 18, pp. 6175–6176, 2012.

[27] A. Ramesh, K. Blanchet, J. H. J. Ensink, and B. Roberts, “Evidence on the Effectiveness of Water, Sanitation, and Hygiene (WASH) Interventions on Health Outcomes in Humanitarian Crises: A Systematic Review,” *PLoS One*, vol. 10, no. 9, p. e0124688, Sep. 2015.

[28] D. S. Lantagne and T. F. Clasen, “Use of Household Water Treatment and Safe Storage Methods in Acute Emergency Response: Case Study Results from Nepal, Indonesia, Kenya, and Haiti,” *Environ. Sci. Technol.*, vol. 46, no. 20, pp. 11352–11360, Sep. 2012.

[29] D. S. Lantagne, R. Quick, and E. D. Mintz, “Household water treatment and safe storage options in developing countries: a review of current implementation practices,” *Wilson Quarterly, Woodrow Wilson Int. Cent. Sch. Environ. Chang. Secur. Progr.*, vol. 99, no. 11, 2006.

[30] H. White, “Theory-Based Impact Evaluation□: Principles and Practice,” no. June, 2009.

[31] T. M. Yates, E. Armitage, L. V Lehmann, A. J. Branz, and D. S. Lantagne, “Effectiveness of Chlorine Dispensers in Emergencies: Case Study Results from Haiti, Sierra Leone, Democratic Republic of Congo, and Senegal,” *Environ. Sci. Technol.*, vol. 49, no. 8, pp. 5115–5122, Apr. 2015.

[32] B. A. Clarke and A. Steele, “Water treatment systems for relief agencies: The on-going search for the ‘Silver Bullet,’” *Desalination*, vol. 248, no. 1–3, pp. 64–71, Nov. 2009.

[33] R. Laxminarayan, J. Chow, and S. Shahid-Salles, “Intervention Cost-Effectiveness: Overview of Main Messages,” in *Disease Control Priorities in Developing Countries*, 2nd ed., D. Jamison, J. Breman, A. Measham, G. Alleyne, M. Claeson, D. Evans, P. Jha, A. Mills, and P. Musgrove, Eds. Washington (DC), 2006.

[34] S. Cairncross and V. Valdmanis, “Water Supply, Sanitation, and Hygiene Promotion,” in *Disease Control Priorities in Developing Countries*, 2nd ed., D. Jamison, J. Breman, A. Measham, G. Alleyne, M. Claeson, D. Evans, P. Jha, A. Mills, and P. Musgrove, Eds. Washington (DC): World Bank, 2006.

[35] “Disease outbreaks,” *The World Health Organization*, 2015. .

[36] “Disease outbreaks by year,” *The World Health Organization*, 2015. [Online]. Available: http://www.who.int/csr/don/archive/year/en/. [Accessed: 14-Sep-2015].

[37] “GIDEON,” 2015. [Online]. Available: http://www.gideononline.com/. [Accessed: 14-Sep-2015].

[38] M. A. Connolly, *Communicable disease control in emergencies: a field manual*. World health organization, 2005.

[39] J. Higgins, R. Churchill, T. Lasserson, J. Chandler, and D. Tovey, “Methodological standards for the conduct of new Cochrane Intervention Reviews (Version 2.3),” 2013.

[40] S. Baird, F. H. G. Ferreira, B. Özler, and M. Woolcock, “Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review,” *Campbell Collab.*, vol. 8, p. 124, 2013.

[41] L. Spencer, J. Ritchie, J. Lewis, and L. Dillon, “Quality in Qualitative Evaluation: A framework for assessing research evidence,” London, 2003.

[42] “Qualitative Research Checklist,” 2013.

[43] “Economic Evaluation Checklist,” London, 2013.

[44] H. Waddington, B. Snilstveit, J. G. Hombrados, M. Vojtkova, J. Anderson, and H. White, “Protocol: Farmer Field Schools for Improving Farming Practices and Farmer Outcomes in Low- and Middle-income Countries: A Systematic Review,” 2012.

[45] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*. 2009.

[46] S. Chinn, “A simple method for converting an odds ratio to effect size for use in meta-analysis,” *Stat. Med.*, vol. 19, no. 22, pp. 3127–3131, 2000.

[47] H. J. Schünemann, A. D. Oxman, J. P. T. Higgins, G. E. Vist, P. Glasziou, G. H. Guyatt, J. P. T. Higgins, and S. Green, “Presenting results and ‘Summary of findings’ tables,” *Cochrane Handb. Syst. Rev. Interv.*, vol. 5, 2008.

[48] J. D. Sachs, “Macroeconomics and Health: Investing in Health for Economic Devleopment,” Geneva, Switzerland, 2001.

[49] “A Manual fo Economic Assessment of Drinking-Water Interventions,” Geneva, Switzerland, 2012.

[50] “Systematic Review Checklist,” 2013.

[51] “11 Questions to Help You Make Sense of a Case Control Study.”

[52] “Tool to Assess Risk of Bias in Cohort Studies.”

[53] M. Bhandari and K. Chan, “Three-minute critical appraisal of a case series article,” *Indian J. Orthop.*, vol. 45, no. 2, p. 103, 2011.

[54] J. P. Higgins and S. Green, Eds., *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons, Ltd, 2008.

[55] “GIDEON,” 2015. .

7 REVIEW TEAM

Name:	**Dr. Daniele Lantagne**
Title:	Assistant Professor
Affiliation:	Tufts University
Address:	Medford, MA, United States
Email:	daniele.lantagne@tufts.edu

Name:	**Travis Yates**
Title:	Research Assistant
Affiliation:	Tufts University
Address:	Medford, MA, United States
Email:	travis.yates@tufts.edu

Name:	**Dr. Myriam Leandre Joseph**
Title:	Physician / Consultant
Affiliation:	Independent
Address:	Port au Prince, Haiti
Email:	mimileandre@hotmail.com

Name:	**Jelena Vujcic**
Title:	Research Scientist / Consultant
Affiliation:	Independent
Address:	Surf City, NC United States
Email:	jelenavujcic@gmail.com

8 DECLARATIONS OF INTEREST

We are not aware of any conflicts of interest that would affect the methods or results presented herein. The research team is also carrying out a separate systematic review on emergency hygiene interventions in emergencies with the International Initiative for Impact Evaluation (3ie). Methodologies and timelines are aligned to benefit and streamline both reviews. Payment was not duplicated, but supplemented for the additional work, mostly in reporting. Ideally, both reviews will complement each other and provide separate publications.

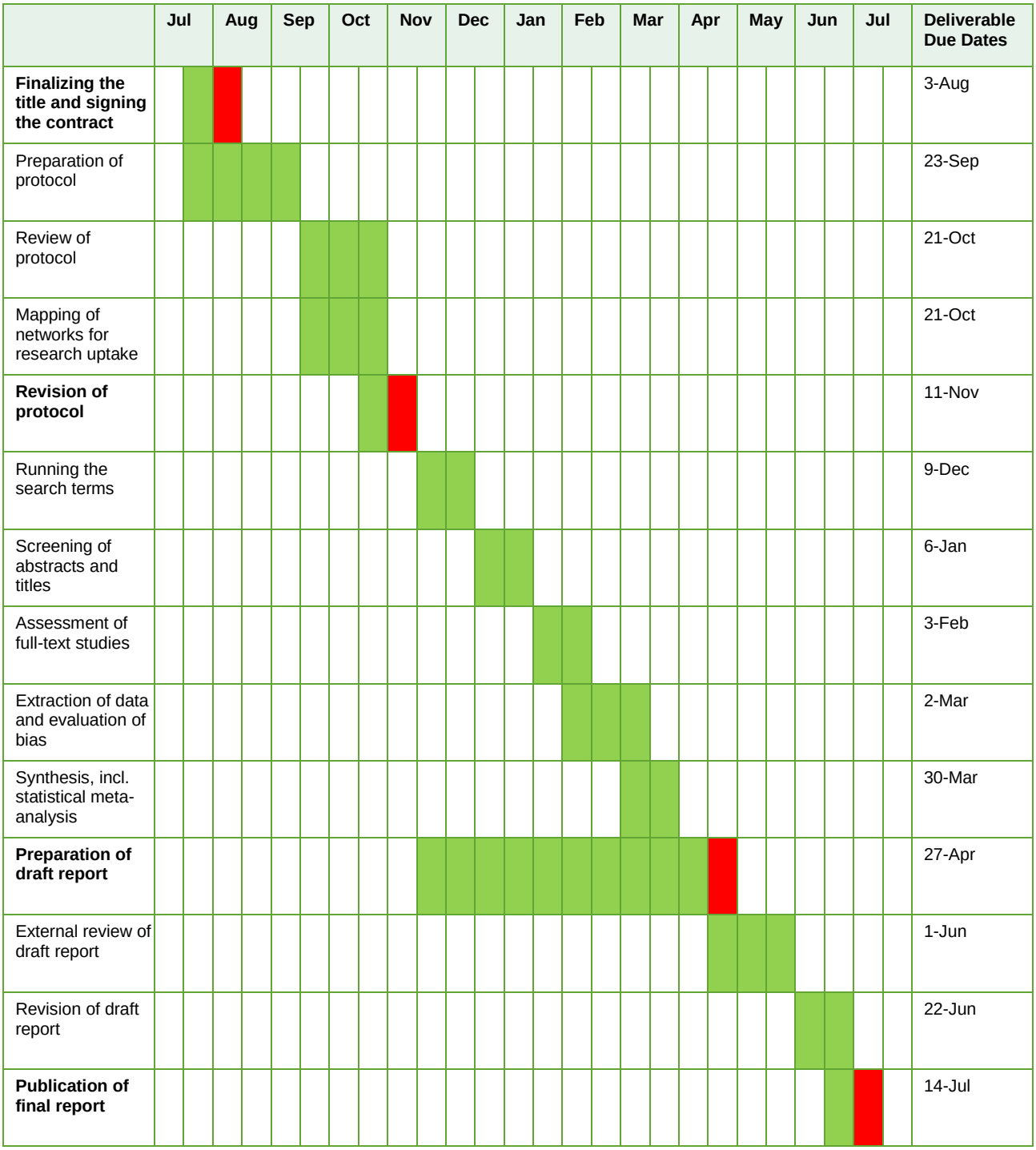
9 ROLES AND RESPONSIBILITIES

The development of the protocol was primarily carried out by Travis Yates with guidance from Daniele Lantagne and contributions from Myriam Leandre Joseph and Jelena Vujcic. Inclusion of manuscripts will be managed by Travis Yates with extensive collaboration by Daniele Lantagne, Myriam Leandre Joseph, and Jelena Vujcic. Data extraction and analysis will be done by Travis Yates and Daniele Lantagne. Final report writing will be led by Travis Yates with input from Daniele Lantagne, Myriam Leandre Joseph, and Jelena Vujcic.

10 PRELIMINARY TIMEFRAME

Start date: 15 July, 2015

End date: 14 July, 2016



	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Deliverable Due Dates
Finalizing the title and signing the contract														3-Aug
Preparation of protocol														23-Sep
Review of protocol														21-Oct
Mapping of networks for research uptake														21-Oct
Revision of protocol														11-Nov
Running the search terms														9-Dec
Screening of abstracts and titles														6-Jan
Assessment of full-text studies														3-Feb
Extraction of data and evaluation of bias														2-Mar
Synthesis, incl. statistical meta-analysis														30-Mar
Preparation of draft report														27-Apr
External review of draft report														1-Jun
Revision of draft report														22-Jun
Publication of final report														14-Jul

11 PLANS FOR UPDATING THE REVIEW

The anticipated limited amount of published data indicates slow cycles of new data. Authors will remain up to date in newly published literature and maintain contacts established through this research for unpublished reports. On March 15, we will re-run electronic searches in peer-reviewed databases to ensure the most relevant data is included in our analysis.

APPENDIX A: DATA COLLECTION VARIABLES

General Information	
First Author	Surname
Year of Publication	(YYYY)
Publication Type	Journal Article Working Paper Book Unpublished Peer Reviewed Unpublished Non-peer Reviewed UN Report (Distributed) NGO Report (Distributed) Other Agency (Distributed) UN Report (non-Distributed) NGO Report (non-Distributed) Other Agency (non-Distributed)
Funder of Intervention	CDC USAid OFDA Unicef UNHCR WHO BMGF HIF DFID ECHO Private Funds Manufacturer Local Government Other Not Reported
Author Affiliation	Employee of intervening body Non-employee of intervening body Consultant Not Reported

Intervention Design	
Implementer (primary agency who received majority of original funds)	International NGO Local NGO UN agency / IFRC / ICRC / IOM Local government Military Other
Intervention Partner	Direct with no local partner Direct and with local partner Indirect with local partner
Target Group	Outbreak-affected Refugee IDP Men Women Children (<5) School age children (5-18 years) Elderly General Population Not Reported
PROGRESS-Plus	Place of Residence Ethnicity Occupation Gender Religion Education Social Capital Socio-economic position Age Disability Sexual orientation Other vulnerable groups
Intervention **(Multiple Answer)**	1) Increasing water access 2) Source-based water treatment options 3) Distribution of household water treatment technologies 4) Promotion of hand hygiene at critical times 5) Distribution of soap and/or hygiene kits 6) Environmental hygiene interventions 7) Installation of temporary or permanent latrines 8) Distribution and management of latrine alternatives

Intervention Design (continued)		
Distribution Component	Yes / No / Unclear	
	if yes to above question **(Multiple Answer)**	Soap Bucket/jerrycan Personal hygiene items Household cleaning Water filter HWT items Cooking supplies Other NFI or CFI materials
Education (Promotion or Behavior change) Component	Yes / No / Unclear	
	if yes to above question **(Multiple Answer)**	Community Household School Radio Other Combination
Promoter Paid	Yes / No / Unclear	
Reference to climate change or climate adaptation	Yes / No / Unclear	
Complementary Programs to WASH Intervention **(Multiple Answer)**	No Yes – Health Yes – Nutrition Yes – Shelter Yes - other	

Timing	
Intervention Period	(MM/YY – MM/YY)
Time from Onset of Outbreak	# of months
Length of Intervention	# of months
Continuation of Intervention Beyond Initial	Yes / No / Unclear

Context	
Global Assessment	Yes / No
Multi-country	Yes / No
Country	Specific country/countries N/A
Region	Sub-Saharan Africa Middle East and North Africa Central Asia South Asia East Asia and Pacific Latin America Caribbean and South America non-LMIC
Outbreak Type	Cholera Typhoid Hepatitis E Respiratory Dysentery Diarrhea Influenza Cryptosporidium Schistosomiasis Ebola Malaria Other
Recurrence	Disease new to area Endemic New disease 2x baseline Spike in cases
Intervention Goal	Prevention Control Both Unclear
Setting	Urban / Rural / Peri-urban
Displacement	Yes / No / Unclear
Camp Setting	Yes / No / Unclear

Study Design				
Study Type (Multiple Answer – economic or mixed methods)	Quantitative RCT / quasi-RCT Case-control Cohort Cross-sectional Non-experimental	Mixed-Methods	Qualitative	Economic
Microbiological testing	Yes / No / Unclear			
Comparison Groups	Yes / No / Unclear			
Purpose of Manuscript	Baseline Intermediate Final Impact Rapid assessment Annual study Global assessment Unclear			
Method of Allocating Groups	Random / Systematic / None / Not Applicable			
Sample Size				
Sample Attrition	Yes / No / Minimal			
Contamination From other interventions	Yes / No / Minimal			

Quantitative Study Quality (Appendix F)	
Selection Bias and Confounding	High Risk / Low Risk / Unclear
Spill-over and Contamination	High Risk / Low Risk / Unclear
Incomplete Outcome	High Risk / Low Risk / Unclear
Selective Reporting	High Risk / Low Risk / Unclear
Other Biases	High Risk / Low Risk / Unclear

Qualitative Study Quality (Appendix F)	
Sample design/target selection of cases/documents?	High Risk / Low Risk / Unclear
Basis of evaluative appraisal?	High Risk / Low Risk / Unclear
How well was the data collection carried out?	High Risk / Low Risk / Unclear
Clarity in reporting and findings?	High Risk / Low Risk / Unclear

Outcomes and Impact		
	Effect Estimation	
	Unadjusted	**Adjusted**
Use of Service		
Economic Impact/Data		
Disease Impact (Morbidity, Mortality, Prevalence, Incidence)		
Non-Health Related Outcomes		
Environmental Impact (Climate Change)		
Additional Context Information Not Captured in Other Categories		

APPENDIX B: ANTICIPATED COMPARISONS

Geography	LMIC Region Fragile States Index Continent
Population	Gender Age Refugee/IDP/ local population
Context	Disease type Additional emergency type Complimentary programming Intervention type Cost-effectiveness New disease to area / endemic
Timing	Time since onset of outbreak Length of intervention Continuation of intervention
Source	Journal/Agency/Grey Donor Agency type

APPENDIX C: KEYWORDS

General: emergency complex crisis humanitarian aid disaster natural disaster outbreak emergency response cholera Ebola hepatitis E dysentery cryptosporidium schistosomiasis malaria diarrhea diarrhoea waterborne diseases disease burden disease risk disease reduction DALY mortality morbidity prevalence evidence effectiveness cost effectiveness efficacy WASH water water quality water quantity sanitation hygiene low income country middle income country LMIC	**Increasing Water Access:** rehabilitation cleaning source protected unprotected improved unimproved tankering **Source-based treatment:** chlorine alum Dispenser HTH well chlorination bucket chlorination pot chlorination **HWT:** PUR aquatab bottled water SwS safe water system chlorine solution HTH sodis filter alum flocculation chlorine water treatment HWT **Handwashing promotion:** hygiene handwashing hand-washing promotion community health worker health worker promoter	**Distribution of soap/hygiene kit:** soap hygiene kit distribution NFI non-food item CRI core relief item **Environmental hygiene:** rubbish collection refuse collection trash collection environmental community plan spraying household cleaning community health worker health worker promoter environmental hygiene **Sanitation facility:** latrine permanent temporary septic tank sanitation **Latrine alternatives:** pee-poo bags port-a-potties port-a-john

APPENDIX D: LIST OF WEBSITES AND ORGANIZATIONS FOR ELECTRONIC SEARCHES

Agency reports and grey literature will be an important data source; thus, we have listed known agencies and websites that are likely to have manuscripts relevant for our review. In situations where websites do not have a searchable database or listed publications, direct solicitation of contacts from the organization will be made.

Type of Organization	Name	Website
UN Agencies	Unicef	http://data.unicef.org/
	WHO	http://www.who.int/gho/database/en/
	UNHCR	http://www.refworld.org/publisher,UNHCR,RESEARCH,,,0.html
	OCHA	https://www.humanitarianresponse.info/en/applications/tools/category/document-repository
International Bodies	International Committee of the Red Cross Red Crescent (ICRC)	https://www.icrc.org/eng/resources/library-research-service/
	International Federation of the Red Cross Red Crescent (IFRC)	http://www.ifrc.org/en/publications-and-reports/evaluations/
	International Organization for Migration (IOM)	http://publications.iom.int/bookstore/index.php?main_page=index&language=en
Development Banks	World Bank	http://data.worldbank.org/
	Asian Development Bank	http://www.adb.org/data/main
	African Development Bank	http://www.afdb.org/en/knowledge/statistics/data-portal/ and http://www.afdb.org/en/knowledge/statistics/open-data-for-africa/
Research Groups	Humanitarian Innovation Fund (HIF)	http://www.elrha.org/hif/innovation-resource-hub/
	EM-DAT The International Disaster Database	http://www.emdat.be/database
	ELRHA	http://www.elrha.org/
	3ie	http://www.3ieimpact.org/evidence/systematic-reviews/ and http://www.3ieimpact.org/en/evidence/impact-evaluations/impact-evaluation-repository/
	Cochrane Collaboration	http://community.cochrane.org/editorial-and-publishing-policy-resource/cochrane-database-systematic-reviews-cdsr
Government Bodies	USAid	http://www.usaid.gov/data
	OFDA	See EM-DAT
	DFID	http://r4d.dfid.gov.uk/
	ECHO	https://euaidexplorer.ec.europa.eu/SearchPageAction.do
	CDC	http://www.cdc.gov/surveillancepractice/data.html

Type of Organization	Name	Website
International Networks	WASH Cluster email list	Personally maintained list
	WASHPlus email list	Personally maintained list
	RedR	http://www.redr.org.uk/
	reliefweb	http://reliefweb.int/topics/wash
	Emergency Environmental Health Forum	Personally maintained list
	ODI	http://www.odi.org/search/site/data
	Humanitarian Practice Network	http://www.odihpn.org/hpn-resources
	Humanitarian Policy Group	Part of ODI
	CDAC Network	http://www.cdacnetwork.org/tools-and-resources/
	Humanitarian Data Exchange	https://data.hdx.rwlabs.org/
NGO	Action Against Hunger (ACF)	http://www.actionagainsthunger.org/technical-surveys/list
	Care International	http://www.care.org/
	International Rescue Committee (IRC)	http://www.rescue.org/
	Oxfam	http://www.oxfam.org.uk/
	Doctors Without Borders (MSF)	http://www.msf.org/reports
	Save the Children	http://www.savethechildren.org/site/c.8rKLIXMGIpI4E/b.6153061/k.7E4A/Publications_and_Reports.htm
	Norwegian Refugee Council (NRC)	http://www.nrc.no/?aid=9137113
	Danish Refugee Council (DRC)	http://drc.dk/home/
	Samaritan's Purse	http://www.samaritanspurse.org/
	Medair	http://relief.medair.org/en/
	World Vision	http://www.worldvision.org/
	Catholic Relief Services	http://www.crs.org/publications/
	PATH	http://www.path.org/publications/list.php

APPENDIX E: SCREENING CHECKLISTS

Screening checklists are intended to help the reviewer identify key aspects of a study without a full review. Screening checklists are used at the second of four filters during the abstract assessment. Each of the six study designs has a screening checklist that is described below. Full assessment criteria are in Appendix F.

E1: SYSTEMATIC REVIEW SCREENING CHECKLIST

Adapted from the systematic review checklist from the Critical Appraisal Skills Programme (CASP) [50]. **Note**: systematic reviews are not included in this review except for cross examining the reference list.

Systematic Review Questions	Yes / No / Unclear
1. Do you think the important, relevant studies were included?	
2. Did the review's authors do enough to assess the quality of the included studies?	

E2: EXPERIMENTAL SCREENING CHECKLIST

Questions for the experimental screening questions were adapted from Waddington et al. (2012) protocol [44].

Experimental Study Design Questions	Yes / No / Unclear
1. Was the random allocation appropriate?	
2. Is the sample size adequate for comparisons?	

E3: QUASI-EXPERIMENTAL SCREENING CHECKLIST

The quasi-experimental questions were adapted by Cochrane and CASP evaluation tools for cohort and case-control studies[51], [52].

Quasi-experimental Study Design Questions	Yes / No / Unclear
1. Was the selection of participants clear and appropriate?	
2. Were populations matched or results adjusted for confounding factors?	

E4: NON-EXPERIMENTAL SCREENING CHECKLIST

The non-experimental study questions were adopted from Bhandari and Chan (2011) [53].

Non-experimental Study Design Questions	Yes / No / Unclear
1. Clear study objective/question?	
2. Explicit inclusion and exclusion criteria for study participants?	

E5: QUALITATIVE SCREENING CHECKLIST

The qualitative study screening questions were adapted from CASP "10 questions to help you make sense of qualitative research" and Spencer et al. 2003 "Quality in Qualitative Evaluation: A Framework for assessing research evidence" [41], [42].

Screening Questions	Yes / No / Unclear
1. Is a qualitative methodology appropriate to meet the objectives?	
2. Is the research design defensible?	

E6: ECONOMIC SCREENING CHECKLIST

The screening questions were adopted from CASP 2013 Economic Evaluations Checklist and the Qualitative Research Checklist [42], [43]. This framework with identify economic manuscripts that are qualitative or quantitative.

Screening Questions	Yes / No / Unclear
1. Was a well-defined question posed? • Both costs and consequences considered? • How many options are compared?	
2. Is there a cost per unit or enough information given to calculate?	
3. Is the research design defensible?	

APPENDIX F: QUALITY APPRAISAL CHECKLISTS

The assessment of different study methodologies require appropriate frameworks unique to each design. The assessment tools listed below are intended to help the reviewer assess manuscripts for common biases and internal validity and are separated by quantitative (G1) and qualitative (G2) research methodologies.

F1: QUANTITATIVE APPRAISAL

To determine the risk of bias in quantitative studies, an assessment tool was developed, drawing heavily from Baird et al (2013) which is based on the Cochrane Handbook Risk of Bias Tool[40], [54]. We will assess the risk of bias through five categories: 1) selection and confounding; 2) spillover and contamination; 3) incomplete outcome; 4) selective reporting; and 5) other bias. Each study will be scored across the five categories as 'Low Risk,' 'High Risk,' or 'Unclear.' The overall determination for the risk of bias for that study is assessed with the table below.

Table 5: Risk of Bias Summary

Risk of Bias	'Low Risk' Assessed in Categories
Low Risk	4-5 'Low Risk' Scores
Medium Risk	3 'Low Risk' Scores
High Risk	1-2 'Low Risk' Scores

F1.1.1: Selection bias and confounding

Bias Score	Criteria
Low Risk	• a. A random component in the sequence generation process is described (e.g. Referring to a random number table) and if the unit of allocation is based on a sufficiently large sample size. • b. The unit of allocation was by geographical/social unit, institution, team or professional and allocation was performed on all units at the start of the study; or if the unit of allocation was by beneficiary or group or episode of treatment and there was some form of centralized randomization scheme, an on-site computer system or sealed opaque envelopes were used. • c. If the outcomes are objectively measurable. • d. Baseline characteristics of the study and control/comparisons are reported and overall similar based on t-test or ANOVA for equality of means across groups. • e. if relevant (e.g. Cluster-rcts), authors control for external factors that might confound the impact of the programme (rain, infrastructure, community fixed effects, etc) through regression analysis or other techniques. • f. The attrition and noncompliance rate is below 15%, or the study assesses whether drop-outs are random draws from the sample (e.g. By examining correlation with determinants of outcomes, in both treatment and comparison groups)?
Unclear	• if a) or b) not specified in the paper, c) scores "no" or if d) scores "no" but the authors controlled for the relevant differences through regression analysis.
High Risk	• Otherwise

Quasi-experimental approaches (non-random allocation of the treatment): was the identification method free from any sources of bias or were sources of bias adequately corrected for with an appropriate method of analysis?

F1.1.2: Quasi-Experimental

Score	Criteria
I. Propensity score matching and combination of psm with panel models:	
Unclear	• a. The study matched on either (1) baseline characteristics, (2) time invariant characteristics or (3) endline variables not affected by participation in the programme. • b. The variables used to match are relevant (e.g. Demographic and socio-economic factors) to explain a) participation and b) the outcome and thus there are not evident differences across groups in variables that explain outcomes. • c. Except for kernel matching, the means of the individual covariates are equal for both the treatment and the control group after matching based on t-test for equality of means or ANOVA.
High Risk	• Otherwise
II. Regression discontinuity design	
Low Risk	• a. Allocation is made based on a pre-determined discontinuity blinded to participants or if not blinded, individuals cannot amend the assignment variable. The sample size immediately at both sides of the cut-off point is sufficiently large. • b. The interval for selection of treatment and control group is reasonably small, or authors have weighted the matches on their distance to the cut-off point. • c. the mean of the covariates of the individuals immediately at both sides of the cut-off point (selected sample of participants and non-participants) are overall not statistically different based on t test or ANOVA for equality of means. • d. If relevant (e.g. Clustered studies) and although covariates are balanced, the authors include control for external factors through a regression analysis.
Unclear	• if a) or b is) not specified in the paper or d) scores “no” but authors control for covariate differences across participants and control individuals.
High Risk	• Otherwise
III. Cross sectional regression studies using instrumental variables and Heckman procedures:	
Low Risk if all the following are true	• a. The instrumenting equation is significant at the level of F ≥ 10; if an F test is not reported, the author reports and assesses whether the Rsquared (goodness of fit) of the participation equation is sufficient for appropriate identification • b. For instrumental variables, the identifying instruments are individually significant (p≤0.01); for Heckman models, the identifiers are reported and significant (p≤0.05) • c. For generalised IV estimation, if at least two instruments are used, the study includes and reports an overidentifying test (p≤0.05 is required to reject the null hypothesis) • d. The study qualitatively assesses the exogeneity of the instrument/ identifier (both externally as well as why the variable should not enter by itself in the outcome equation); only score yes when the instrument is exogenously generated: e.g. natural experiment or random assignment of participants to the control and treatment groups. If instrument is the random assignment of the treatment, the systematic reviewer should assess the quality and success of the randomisation (e.g. see section on RCTs). • e. The study includes relevant control for confounding, and none of the controls is likely affected by participation.
Unclear	• if d) scores “no” and c) scores “yes”.
High Risk	Otherwise

F1.1.2: Quasi-Experimental (continued)

Score	Criteria
IV. Cross sectional regression studies using OLS or maximum likelihood models including logit and probit models.	
Unclear if all the following are true	• The covariates distribution are balanced across groups • The authors control for a comprehensive set of confounders that may be correlated with both participation and explain outcomes (e.g. demographic and socio-economic factors at individual and community level) and thus, it is not evident the existence of unobservable characteristics that could be correlated with participation and affect the outcome. • The authors use proxies to control for the presence of unobservable confounders driving both participation and outcomes. • Participation does not have a causal impact in any of the controls.
High Risk	• Otherwise
V. Panel data models (controlled before-after, difference in difference multivariate regressions):	
Unclear if all the following are true	• The authors use a difference in difference multivariate estimation method or fixed effects models. • The author control for a comprehensive set of time-variant characteristics (e.g. the study includes adequate controls for confounding and thus, it is not evident the existence of time-variant unobservable characteristic that could be correlated with participation and affect the outcome) • The attrition and noncompliance rate is below 10%, or the study assesses whether drop-outs are random draws from the sample (e.g. by examining correlation with determinants of outcomes, in both treatment comparison group)?
High Risk	• Otherwise

F1.1.3: Non-Experimental

Score	Criteria
Non-experimental studies	
Unclear	• Mixed methods – individual components of mixed-methods research need to be assessed independently and scored. It is possible that quantitative data from a mixed method study scores a ‘high bias’ and qualitative scores a ‘low bias’ or vice versa.
High Risk	• Case reports • Case series • Uncontrolled before-after • Correlation research • Single variable research – no control or comparison group

F1.2 Spillovers and contamination

Score	Criteria
Was the study adequately protected against spillovers, cross-overs and contamination?	
Yes	• The intervention is unlikely to spillover to comparisons (e.g. Participants and non-participants are geographically and/or socially separated from one another and general equilibrium effects are not likely) and that the treatment and comparisons are isolated from other interventions which might explain changes in outcomes.
No	• Allocation was at the individual level and there are likely spillovers within households and communities which are not controlled for, or • Other interventions likely to affect outcomes operating at the same time in either group.
Unclear	• Spillovers and contamination are not addressed clearly

F1.3 Incomplete Outcome Data

Attrition bias due to amount, nature or handling of incomplete outcome data	
Score	**Criteria**
Low risk	• No missing outcome data; • Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • Missing data have been imputed using appropriate methods • Authors use 'common' methods of estimation (i.e. Credible analysis method to deal with attribution given the data available). Additionally, specific methods of analysis should answer positively the following questions: • For RCTs, if randomisation clearly described and achieved, e.g. Comparison of treatment and control on all appropriate observables prior to selection. • For PSM, if (a) for failure to match over 10% of participants, sensitivity analysis is used to re-estimate results using different matching methods (kernel matching techniques); (b) for matching with replacement, there is not any observation in the control group that is matched with a large number of observations in the treatment group; (c) authors report the results of rosenbaum test for hidden bias which suggest that the results are not sensitive to the existence of hidden bias. • For IV and Heckman models, if (a) the author tests and reports the results of a hausman test for exogeneity ($p \leq 0.05$ is required to reject the null hypothesis of exogeneity); (b) the study describes clearly and justifies the exogeneity of the instrumental variable(s)/identifier used (iv and heckman); (c) the value of the selectivity correction term (rho) is significantly different from 0 ($p<0.05$) (heckman approach). • d. For regression analysis, if authors carried out a hausmann test with a valid instrument and the authors cannot reject the null of exogeneity of the treatment variable at the 90% confidence.

Attrition bias due to amount, nature or handling of incomplete outcome data	
Score	**Criteria**
High Risk	• Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size; • 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; • Potentially inappropriate application of simple imputation.
Unclear	• Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided); • The study did not address this outcome

F1.4 Selective Reporting

Reporting bias due to selective outcome reporting	
Score	**Criteria**
Low risk	• The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
High Risk	• Not all of the study's pre-specified primary outcomes have been reported; • One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; • One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; • The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Unclear	• Insufficient information to permit judgement of 'Low risk' or 'High risk'. **It is likely that the majority of studies will fall into this category.**

F1.5 Other Bias

Bias due to problems not covered elsewhere in the table.	
Score	**Criteria**
Low risk	• The study appears to be free of other sources of bias.
High Risk	• Data was collected by self-reporting from the beneficiary • Blinding of the outcome may not have been controlled – lack of blinding. • Alternation or rotatation of enrolment, also concealment by date of birth or case number or any other explicitly unconcealed procedure. • Other potential threats to validity are present, and note these below (e.g. Coherence of results, data on the baseline collected retrospectively, information is collected using an inappropriate instrument or a different instrument/at different time/after different follow up period in the control and in the treatment group). • Had a potential source of bias related to the specific study design used; or • Has been claimed to have been fraudulent; or • Had some other problem.
Unclear	• Insufficient information to assess whether an important risk of bias exists; or • Insufficient rationale or evidence that an identified problem will introduce bias.

F2: QUALITATIVE APPRAISAL

The qualitative assessment has been adapted from Spencer et al. 2003 "Quality in Qualitative Evaluation: A Framework for assessing research evidence" [41], [42]. The quality assessment is evaluated on four appraisal questions. There is no clearly objective rule for determining bias among qualitative studies. The guiding questions will be used by the research team to help establish core research questions that should be evaluated; however, professional judgment is necessary to make the assessment. Qualitative experts are on the research team and advisory board to ensure rigorous standards, consistency, and transparency.

Each study will be scored across the four appraisal questions categories as 'Low Risk,' 'High Risk,' or 'Unclear.' The overall determination for the risk of bias for that study is assessed with the table below.

Table 6: Risk of Bias Summary

Risk of Bias	'Low Risk' Assessed in Categories
Low Risk	3 or more 'Low Risk' Score
Medium Risk	2 'Low Risk' Scores
High Risk	1 or less 'Low Risk' Score

Appraisal Questions	Guiding Questions	Low Bias / High Bias / Unclear
1. How well defended is the sample design/ target selection of cases/ documents?	• Description of study locations/areas and how and why chosen • Description of population of interest and how sample selection relates to it (e.g. typical, extreme case, diverse constituencies etc.) • Rationale for basis of selection of target sample/settings/documents (e.g. characteristics/features of target sample/settings/documents, basis for inclusions and exclusions, discussion of sample size/number of cases/setting selected etc.)	
2. How clear is the basis of evaluative appraisal?	• Discussion of how assessments of effectiveness/evaluative judgments have been reached (i.e. whose judgments are they and on what basis have they been reached?) • Description of any formalized appraisal criteria used, when generated and how and by whom they have been applied • Discussion of any unintended consequences of intervention, their impact and why they arose	
3. How well was the data collection carried out?	• Who conducted data collection? • Were there procedures/documents used for collection/recording (Audio or video recording) • Examination of origins/influences on opposing or differing positions	
4. Is there clarity in reporting and findings?	• How clear and coherent is the reporting? • Demonstrates link to aims of study/research questions? • How clear are the assumptions/ theoretical perspectives/values/richness of data that have shaped the form and output of the evaluation?	

F3: ECONOMIC APPRAISAL

The economic assessment has been adopted from CASP Economic Evaluation Checklist (2013) [43]. If two of the three questions are 'high risk' or 'unclear' the study is considered high risk overall. One 'high risk' from the three categories and the overall assessment is 'medium risk,' otherwise, 'low risk.'

Appraisal Questions	**High Risk / Low Risk / Unclear**
1. Were all important and relevant resources required and health outcome costs for each alternative identified, measured in appropriate units and valued credibly? Consider how realistic are they and how they were derived?	
2. Were sensitivity and incremental analyses preformed? Consider changing the estimate of the variable does this change the result of the economic evaluation?	
3. Are results transferable to other contexts? Consider costs and program being translatable to other settings.	

Published by Oxfam GB for Oxfam International under 978-0-85598-716-9 in April 2016.
Oxfam GB, Oxfam House, John Smith Drive, Cowley, Oxford, OX4 2JY, UK.

OXFAM

Oxfam is an international confederation of 20 organisations networked together in more than 90 countries, as part of a global movement for change, to build a future free from the injustice of poverty. Please write to any of the agencies for further information, or visit www.oxfam.org.