

# PROTOCOL: Water, sanitation and hygiene for reducing childhood mortality in low- and middle-income countries

Hugh Sharma Waddington<sup>1</sup> | Sandy Cairncross<sup>2</sup>

<sup>1</sup>London School of Hygiene and Tropical Medicine, London International Development Centre, London, UK

<sup>2</sup>London School of Hygiene and Tropical Medicine, London, UK

## Correspondence

Hugh Sharma Waddington, London School of Hygiene and Tropical Medicine, London International Development Centre, WC1A 2NS, London, UK.

Email: [Hugh.waddington@lids.ac.uk](mailto:Hugh.waddington@lids.ac.uk)

## Abstract

Respiratory tract infections and diarrhoea are the two biggest killers of children in low income contexts. They are closely related to access to, and use of improved water, sanitation and hygiene (WASH). However, there is no high quality systematic review that quantifies the effect of WASH improvements on childhood mortality. Existing systematic reviews of WASH improvements measure effects on morbidity, under the (often implicit) assumption that morbidity is closely correlated with mortality. This is at least partly because the impact evaluations on which they are based are only designed to detect changes in morbidity with statistical precision, whereas mortality is a relatively rare outcome. The proposed review will address this evidence synthesis gap, using the greater statistical power of meta-analysis to pool findings across studies.

## 1 | BACKGROUND

### 1.1 | One thousand children died today because of diarrhoea

Water, sanitation and hygiene (WASH) are fundamentally important for human life, health and happiness. Maslow (1943) proposed a hierarchy of goals for human life in the following order: “physiological, safety, love, esteem, and self-actualization” basic needs. The physiological needs relate to healthy regulation of the human body’s metabolism via sufficient access to air, water, nutrition, warmth, rest (including sleep) and the means to excrete. Safety was placed just above physiological needs, and linked specifically to safety from illness and pain in childhood, as well as from “wild animals” and “assault” throughout the life-course. It is quite difficult to over-emphasise the contribution of sufficient water, sanitation and hygiene to ensuring basic needs are met.

Yet, according to the Joint Monitoring Programme (JMP), two billion people do not have safe, readily available water at home, and 4.5 billion lack access to safely managed sanitation services (WHO/UNICEF, 2019). In sub-Saharan Africa, 400 million people have to use

surface water or improved water sources that take more than 30 min to reach, queue for and return home. Of the 1.4 billion people worldwide who defecate in the open or use unimproved or shared sanitation facilities, half a billion live in South Asia (around 400 million in India) and another half billion are in sub-Saharan Africa.

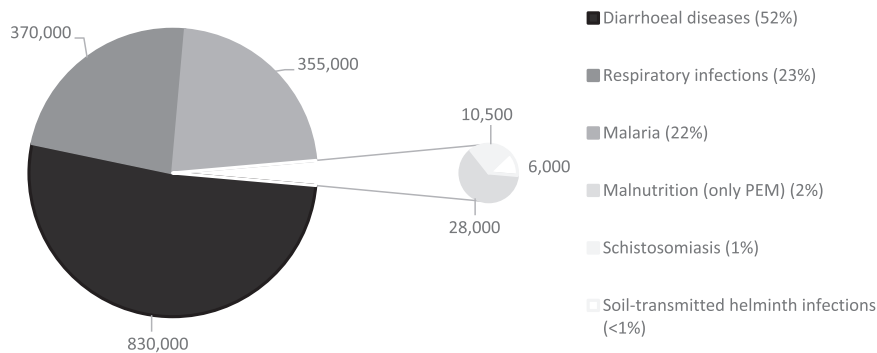
Inadequate WASH can contribute to the outbreak and chronic presence of preventable infections like pneumonia and diarrhoeal disease, which are the two biggest killers of children globally (Liu et al., 2012).<sup>1</sup> Enteric disease may also cause tropical enteropathy, a subclinical disorder where the lining of the gut wall is damaged by repeated bouts of infection until it is unable to absorb nutrients adequately (Humphreys, 2009; Shiffman et al., 1978). Chronic high enteric infection rates are among the leading causes of under-nutrition and death in children in developing countries (Cairncross et al., 2014).

Water-related diseases are responsible for an estimated 21% of the global disease burden (Black et al., 2010). According to recent Global Burden of Disease (GBD) estimates (Prüss-Ustün et al., 2019),

<sup>1</sup>Hygiene and water supply are also likely to be key blocks to the transmission of coronavirus disease 2019 (COVID-19), a type of acute respiratory tract infection (Howard et al., 2020).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Campbell Systematic Reviews* published by John Wiley & Sons Ltd on behalf of The Campbell Collaboration



**FIGURE 1** Estimated annual global deaths due to inadequate WASH. PEM protein energy management. Source: Chirgwin et al. (submitted).

inadequate WASH is associated with 1.6 million deaths per year, due to diarrhoea, acute respiratory infection, malnutrition due to protein energy management and, as a result of water mismanagement, malaria (Figure 1). Diarrhoea alone kills 850,000 people every year, 300,000 of whom are children aged under 5 (Prüss-Ustün et al., 2019). This is equivalent to a line of dead children from London to Cardiff every year, every death being a personal tragedy (White, 2004).

Parasitic worm infections, associated with inadequate sanitation (e.g., *Ascaris*, *Trichuris* and hookworm infections), are responsible for 39 million disability-adjusted life years (DALYs), equivalent to the global burden of mortality for malaria and tuberculosis combined (Stephenson et al., 2000; see also Ziegelbauer et al., 2012). Trachoma, a water-washed eye infection causing blindness, spread by the *Musca sorbens* fly which breeds in human excrement, affects an estimated 146 million people worldwide (Freeman et al., 2017; Rabiou et al., 2012). Water supply changes may also affect rates of arsenic poisoning due to groundwater consumption, which can cause nutritional deficiency, cancer and death (Dar & Khan, 2011; Jones-Hughes et al., 2013).

There may also be important externalities from private consumption of improved WASH services through environmental health spillovers (Barreto et al., 2007; Duflo et al., 2015; Root, 2001; Spears, 2013). These operate in the private domain (household and yard) and public domains (places of work, education, commerce, recreation, street and fields) (Cairncross et al., 1996). For example, the World Bank (2008) estimated environmental costs of poor sanitation at 2% of GDP in South Asia (Cambodia, Indonesia, the Philippines and Vietnam). Poor access and use of WASH in places with high population density, may explain why some countries, particularly in South Asia, have worse child malnutrition outcomes than their income levels alone would predict (Spears, 2013).

While all suffer loss of dignity from open defecation and drudgery from water collection, women and girls suffer particularly. Women do the majority of water carrying when households lack access to an improved water source in Africa and Asia (Sorenson et al., 2011). Originally, McSweeney (1979) had reported that the burden of time spent on domestic chores in Burkina Faso started in a girl's childhood, was around 7–8 h/day by age 9 (double that of boys of similar age) and women and girls were responsible for all of the water collection. Feachem et al. (1978) estimated that 96% of water

collections in Lesotho were made by women and girls. Cairncross and Cliff (1987) reported time savings associated with water supply improvements for women in Mozambique, which were put to other household activities (food preparation and childcare), suggesting a possible mechanism through which WASH impacts on nutrition, and therefore possibly child survival (see also Dangour et al., 2013). Women and girls still do most water collection in 24 sub-Saharan Africa countries (Graham et al., 2016), risking becoming pedestrian road casualties, and risking attack and assault by “pests and perverts” (Campbell et al., 2015). For example, Cairncross and Cliff (1987) found in northern Mozambique that, when the functioning village standpipe broke down, women were forced to rely on traditional sources. The choice included a water source 8 km away, taking between 4 and 7 h (travel time and queueing) for the return journey, or one 4 km away, where “[a] few women spent the night... despite the danger of lions, waiting for water to appear in the holes dug for that purpose” (p. 51).

Women and girls can also be put in danger when they have to wait until after dark to urinate or defecate with privacy (Sommer et al., 2014; Sorenson et al., 2011). For example, studies in Kenya (Winter & Barchi, 2016) and India (Jadhav et al., 2016) found that women who openly defaecated were more likely to experience nonpartner sexual and/or physical violence; in India, which compared women who openly defaecated with those with a private toilet, the difference was 200%. There may also be adverse maternal and child health implications due to inadequate WASH services in health facilities and other places of newborn delivery (Benova et al., 2014). Pregnant women and neonates are thought to be a particularly high-risk group because infection and sepsis are major causes of maternal and neonatal mortality (Liu et al., 2012). More generally, disadvantaged groups, such as children, the elderly, women, poor people, immunocompromised people such as those living with human immunodeficiency virus (HIV) and people with disabilities, are less likely to have access to appropriate WASH technologies (whether drinking water supplies of sufficient quantity and quality, means of safe excreta disposal and hygiene practices), and therefore more likely to experience mortality and negative health and socioeconomic consequences.

Where female adults are required to collect the water, which is the majority of cases, older children may be pulled out of the school

to care for younger ones (Koolwal & van de Walle, 2010). Diminished educational attainment, due to children's school enrolment and attendance as well as teacher attendance, as well as delayed entry to the labour market, have implications for employment, life-time wage earnings and poverty (Hutton et al., 2007; Poulos et al., 2006).

Other longer-term economic implications arise due to delayed entry to the labour market, and monetary losses due to costs of medical treatment and aversion costs of treating and storing unclean water or purchasing water from vendors (Bosch et al., 2002; Cairncross & Kinnear, 1992). These costs can be exorbitant for poor households in urban informal settlements (slums) who are unserved by house connections. For example, the costs of vendor supply were estimated at 7–11 times higher than public utility water supply in Nairobi, Kenya, 12–25 times in Dacca, Bangladesh, 28–83 times higher in Karachi, Pakistan, 17–100 times higher in Port-au-Prince, Haiti, and 100 times higher in Nouakchott, Mauritania (Bhatia & Falkenmark, 1993, p. 14). In a study in Khartoum, Sudan, where up to 56% of household income in squatter areas was spent on vendor water (Cairncross & Kinnear, 1992), the income and price elasticities of demand for water were found to be very inelastic (that is, demand is relatively unresponsive to changes in income and price). It was therefore suspected that the poorest households would need to substitute food expenditure to meet water needs, causing malnutrition.

For all of these reasons, inadequate WASH service access and use is likely to support vicious cycles of limited human development and weak economic growth (Ramirez et al., 1998). It is very important, therefore, to understand the likely magnitude of the impacts of WASH interventions on important outcomes, such as mortality, in particular contexts and for particular groups.

## 1.2 | Water, sanitation and hygiene interventions

WASH interventions have several important components to them (Chirgwin et al., submitted) including: the technology that is provided to users (e.g., a child's potty and knowledge about safe excreta disposal); the intervention mechanism used to encourage demand among the target population (e.g., a government subsidy on the potty purchase price and promotional campaign about excreta disposal) or to improve supply (e.g., capacity building for sanitation providers); and the social and physical environment where participants use the technology (e.g., the household and yard).<sup>2</sup>

### 1.2.1 | WASH technologies

The quality of water supply, sanitation and hygiene facilities—that is, the extent to which they are likely to provide drinking water of sufficient quantities for basic needs, enable hygienic hand-washing and food

preparation, and safe removal of excrement from the human environment—is dependent on the type of technology. These are usually grouped into drinking water, sanitation and hygiene ladders (Table 1).<sup>3</sup>

### 1.2.2 | Intervention mechanisms

Mechanisms for providing WASH technologies can be categorised into demand and supply side interventions. Demand side interventions include: behaviour change communication, such as health education and psychosocial “triggering,” for example, social marketing and community-led total sanitation (CLTS); subsidies and microloans for consumers; and legal measures proscribing open defaecation, discharge of contaminated water, or dumping of waste (e.g., Cairncross, 1992). For example, psychosocial triggering aims to promote demand for WASH technology among consumers using directive or participatory methods (De Buck et al., 2017). An example of a directive approach is social marketing, which motivates social change through a combination of product (technology used to meet a need), promotion (to increase desirability and acceptability), place (installation in an appropriate place for users) and price (the cost for users takes into account affordability) (Cairncross, 2004; Evans et al., 2014). These are often implemented at community level such as in schools and health facilities via approaches such as community health clubs to promote demand (Waterkeyn & Cairncross, 2005). Participatory, bottom-up approaches are also being rapidly scaled up, including CLTS. In CLTS, the community is facilitated to discuss how they would like sanitation practices to change, identify problem areas (e.g., “walks of shame”), and use social cohesion and pressure to motivate people to construct latrines and stop practising open defaecation (Kar & Chambers, 2008).

On the supply side, interventions include: direct provision of technology by an external body (e.g., government, NGO); improving operator performance (e.g., institutional reform, capacity building, operator financing, regulation and accountability); privatisation (e.g., Galiani et al., 2005) and nationalisation of service delivery; and promoting small-scale independent provider involvement (e.g., sanitation marketing through microloans and capacity building for providers) (Poulos et al., 2006). WASH technology may be for use in private (household and yard) or public spaces (shared facilities, WASH in health facilities and schools, places of transit, work, commerce, reaction, streets and fields). Measures to improve service provider performance include measures such as enacting and implementing water quality standards (Cairncross et al., 1996), government regulation of private utility providers, and reforms to operator financing (e.g., payment-by-results) (Poulos et al., 2006). Encouraging small-scale independent providers like nonprofits and the private sector (Sansom et al., 2003) may include microloans for WASH service providers and capacity building. As an example of the

<sup>2</sup>A fourth component, not discussed here, is the suitability of the intervention for particular groups including disadvantaged people (e.g., children, pregnant women, elderly and disabled people).

<sup>3</sup>There are also intermediate steps on the sanitation ladder not listed in Table 1; for example, where there is no fixed place of sanitation but some attempt to remove faeces from exposure to others such as “cat sanitation” (Waterkeyn and Cairncross, 2005).

**TABLE 1** Water, sanitation and hygiene technology ladders

	Drinking water	Sanitation	Hygiene
Improved facilities: safely managed	Improved facilities that: <ul style="list-style-type: none"> <li>• are accessible on premises, and</li> <li>• provide water when needed, and</li> <li>• provide water free from contamination.</li> </ul>	Improved facilities where waste products are either: <ul style="list-style-type: none"> <li>• treated and disposed in situ, or</li> <li>• temporarily stored and then emptied and transported to off-site treatment centre, or</li> <li>• transported through sewer with wastewater and treated off-site</li> </ul>	Undefined
Improved facilities: basic	Improved sources that require <30 min round-trip to collect (including queueing time). These include piped supplies: <ul style="list-style-type: none"> <li>• tap water in the dwelling, yard, or plot</li> <li>• public standposts/pipes</li> <li>• And nonpiped supplies: <ul style="list-style-type: none"> <li>• boreholes/tubewells</li> <li>• protected wells and springs</li> <li>• rainwater</li> <li>• packaged water, including bottled water and sachet water</li> <li>• delivered water, including trucks and small carts</li> </ul> </li> </ul>	Improved facilities provided at the household level. These include networked sanitation: <ul style="list-style-type: none"> <li>• flush and pour flush toilets connected to sewers</li> <li>• And on-site sanitation: <ul style="list-style-type: none"> <li>• flush or pour flush toilets connected to septic tanks or pits</li> <li>• pit latrines with slabs</li> <li>• composting toilets, including twin pit latrines and container-based systems</li> </ul> </li> </ul>	Fixed or mobile handwashing facilities with soap and water: <ul style="list-style-type: none"> <li>• handwashing facilities defined as a sink with tap water, buckets with taps, tippy-taps and jugs or basins designated for handwashing</li> <li>• soap includes bar soap, liquid soap, powder detergent and soapy water</li> </ul>
Limited facilities	Improved sources of the above types requiring more than 30 min to collect including queueing time	Improved facilities of the above types shared by two or more households	Handwashing facilities without soap and water (e.g., ash, soil, sand or other handwashing agent)
Unimproved facilities	Nonpiped supplies: <ul style="list-style-type: none"> <li>• unprotected wells and springs</li> </ul>	On-site sanitation or shared facilities of the following types: <ul style="list-style-type: none"> <li>• pit latrines without slabs</li> <li>• hanging latrines</li> <li>• bucket latrines</li> </ul>	Undefined
No facilities	Surface water (e.g., drinking water directly from a river, pond, canal, or stream)	Open defecation (disposal of human faeces in open spaces or with solid waste)	No handwashing facility on premises

Source: WHO/UNICEF (2019); <https://washdata.org/monitoring>.

latter, sanitation marketing aims to increase availability of sanitation technology and maintenance services (such as pit emptying), by training local artisans to produce sanitation products that are suitable for the varying needs of consumers (e.g., Cameron et al., 2013).

Decentralisation, where community representatives are placed in planning, design, implementation and operation of the WASH service provider, is an example of an intervention category that combines supply and demand. For example, community-driven development (CDD) uses a bottom-up approach, block grants with cost sharing and often a component of local institutional strengthening (White et al., 2018).

### 1.2.3 | Place of use

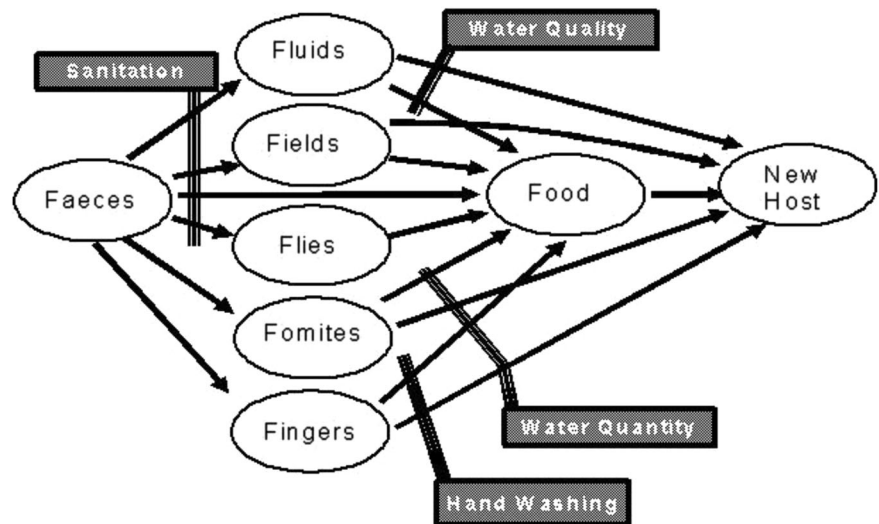
The social and physical environment where participants interact with WASH technology is important for understanding infectious disease

transmission. As noted, Cairncross et al. (1996) distinguished private domain (dwelling and yard) and public domains (community, schools, places of work, commerce and recreation, fields in rural areas and streets in cities) in disease transmission. The importance of the differentiation is in the potential for communicable disease transmission—the greater potential for single cases to cause epidemics in public spaces—and the types of interventions that are needed to combat transmission—the greater focus on infrastructure investment and regulation in public space, and personal hygiene in private spaces (which also depends on infrastructure investment especially water supply).

### 1.3 | The effects that WASH access and use can have

Outcomes of WASH sector interventions can be categorised into five groups: intermediate outcomes relating to WASH access, knowledge,

**FIGURE 2** The “F”-diagram showing faecal-oral disease transmission. Source: Cairncross and Feachem (2018).



attitudes and behaviours (e.g., time use, consumer satisfaction, environmental pathogen contamination); health outcomes due to water-related health infection (e.g., diarrhoeal infection, acute respiratory infections, gastro-intestinal worm infections); other health outcomes, which are largely gendered (musculoskeletal disorder, reproductive tract infection, injury and psychosocial health); nutritional status, relating to water-related disease and carer and children's time use; socioeconomic outcomes (e.g., education and cognitive development, income poverty); and mortality (Chirgin et al., submitted).

Figure 2 shows a theoretical depiction of the direct communication of faeco-oral pathogens between individuals (Wagner & Lanoix, 1958). Later called the “F-diagram” (e.g., Kawata, 1978), it shows the behavioural transmission routes for various water-related diseases from faeces to future hosts via water (fluids), hands (fingers), arthropods (flies), soil (fields) and food. A sixth transmission route has since been identified, “fomites”—that is, objects acting as disease-carrying vectors such as clothes, utensils, toys and furniture (Cairncross & Feachem, 2018). Implicit in the figure are three water-related, faecal-borne disease transmission routes: water-borne diseases transmitted through ingesting infected water, water-washed diseases transmitted through inadequate drinking water supply and hygiene (e.g., cholera, diarrhoeal disease, hepatitis, typhoid), and water-based diseases transmitted by penetrating skin (e.g., schistosomiasis transmitted in water, and ascaris, hookworm and whipworm in contaminated soil) (White et al., 1972, p. 163).

The F-diagram focuses on faecal-borne diseases, but additional water-related infections that are not faeces-related. For example, water-related insect vectors which pass on disease by breeding in water (e.g., chikungunya, dengue, malaria) (Cairncross & Feachem, 2018) are a major source of global mortality (Figure 1).

Figure 2 shows sanitation as a primary barrier to faecal-related disease transmission, when excreta carrying faecal pathogens are eliminated from the environment or human consumption. Primary barriers also include hand washing and water quantity, important for stopping

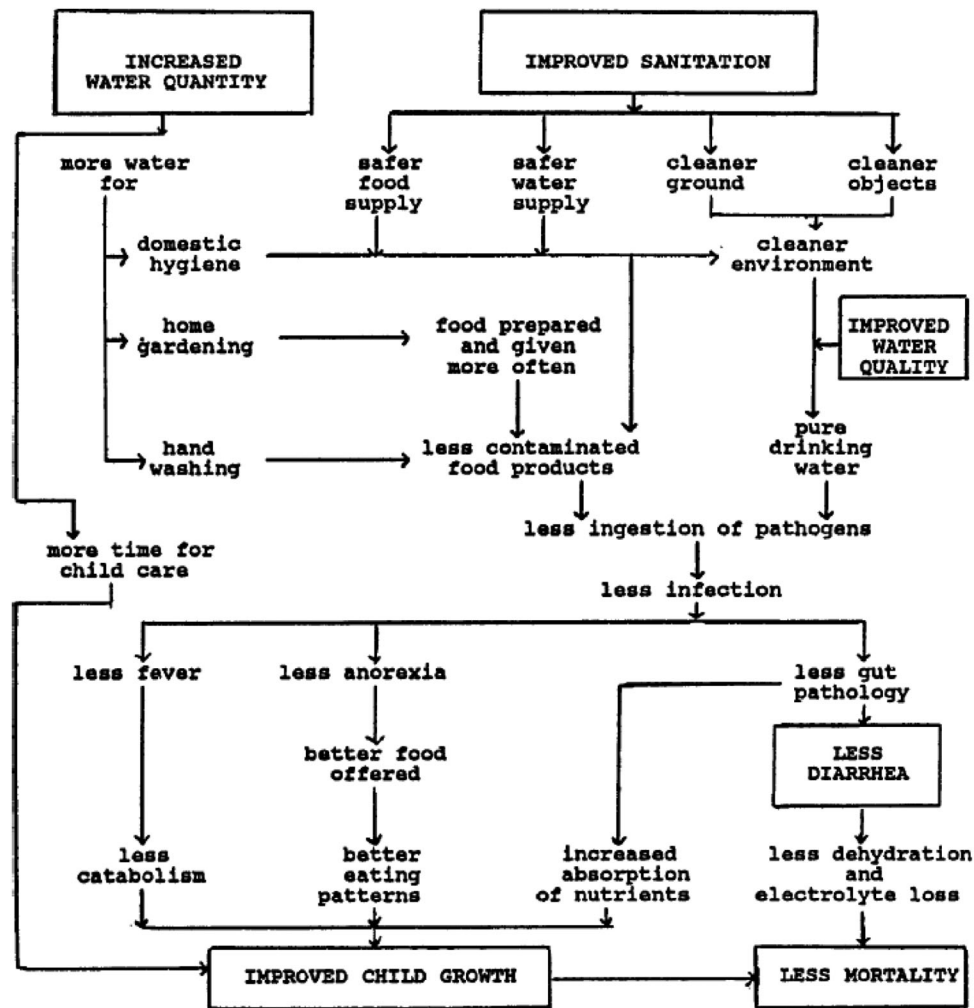
transmission primarily in the domestic domain (fingers and fomites). Due to faecal contamination of drinking water between source and point-of-use (POU), hygienic approaches may be needed to store clean water collected at source, or treat water for contaminants in the household (POU) (Fewtrell & Colford, 2004; Wright et al., 2004). Better access to water supply (quantity) may improve health by reducing contamination in the environment by enabling better personal hygiene (e.g., handwashing) and environmental hygiene (e.g., safe disposal of faeces). The secondary barrier is drinking water quality (Kawata, 1978). Factors such as environmental faecal contamination may prevent impacts from clean drinking water provision being realised due to the amount of time infants and children, who are the most susceptible to diarrhoeal disease, spend on the floor and putting their fingers in their mouths.<sup>4</sup>

Esrey (1987) presented a logic model showing the theoretical relationship between water supply, water treatment, sanitation and hygiene, on the one hand, and diarrhoeal disease, child nutritional status, and survival, on the other (Figure 3). That figure indicates that the routes from water supply and sanitation to survival operate through various intermediate quality of life outcomes relating to better hygiene practices (including hand- and food-hygiene, and “fomites”) and child care, diarrhoeal disease and nutrition.

The figure is highly simplified and excludes underlying assumptions. Links in the causal pathway between interventions and outcomes are not automatic. For example, water treatments may not lead to less faecal contamination if the treatment technology itself is not efficacious in combating parasitic infections (Arnold & Colford, 2007). An example would be chlorination which is not effective against cryptosporidium, a common cause of diarrhoeal morbidity and mortality, especially among immunocompromised

<sup>4</sup>The F-diagram relates to faecal-borne pathogen related disease transmission. Non-infectious waterborne diseases, such as arsenicosis and fluorosis, caused through chemical contamination of water, are increasingly recognised as a source of human morbidity and mortality (Dar & Khan, 2011).





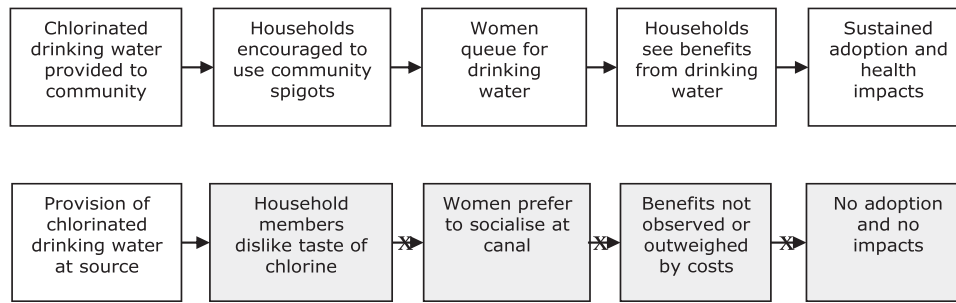
**FIGURE 3** Relationship of improved water, sanitation and hygiene to diarrhoea, child growth and mortality among young children. Source: Esrey (1987).

groups such as those living with HIV/AIDS (Havelaar et al., 2003; cited in Abubakar et al., 2007). And even an efficacious technology may not reduce contamination if used improperly, for example, where insufficient protective agents are applied to treat drinking water, or insufficient time available to purify water before ingestion. In the case of drinking water provided at source, there may be environmental contamination during transport (e.g., use of contaminated storage containers) or poor personal hygiene at POU (e.g., when contaminated hands are put in water storage containers) (Wright et al., 2004). Other factors limiting effectiveness are due to adoption, for example, users may dislike the odour and taste of chlorinated water.

Similarly, providing latrines may not necessarily lead to less open defaecation (Clasen et al., 2010), for various reasons such as the quality of facilities (cleanliness and smell) or concerns from pit owners about the frequency that the pit will need to be emptied. Nor may latrine provision lead to better health and nutrition if open defaecation is still practised by some people in densely populated

areas (Kar & Chambers, 2008). Latrines are not usually designed for or used by children, who may be afraid of going into dark places or of falling into the pit. This may be particularly problematic for reducing environmental contamination because children's excreta are more likely to contain infectious pathogens than adults' (Cairncross & Feachem, 2018), even though they may not be thought dangerous or offensive (Curtis et al., 1995; see also Majorin et al., 2019).

Preventive technologies tend to be adopted more slowly as benefits are difficult to observe (Rogers, 2005). This applies particularly to WASH technologies whose main benefit is to reduce diseases, the prevalence of which may typically be infrequent (or effects unobserved) outside of epidemics. For example, the incidence of diarrhoeal disease among study participants in low- and middle-income countries (L&MICs) was around 10% in one systematic review (Waddington et al., 2009). An average reduction in risk of child diarrhoea by 30%, the typical pooled effect size found in meta-analyses of WASH technology evaluations, would therefore only reduce the number of



**FIGURE 4** Programme theory and practice: public spigots in Egypt. Source: Author drawing on the description contained in Rogers (2005).

diarrhoeal days from 10% to 7% on average, if the measure were based on prevalence.<sup>5</sup> Even a reduction in average risk by 50% for household water filtration, would reduce the typical child diarrhoeal risk from three episodes per year to 1.5 episodes (Clasen et al., 2015). In contrast, where the benefits of a technology are easily observed by those directly affected, such as poor women and children collecting water every day, and hence adoption likely to be rapid where it can be adequately provided, it is more likely that underinvestment in the technology would be explained by systemic undervaluation of the benefits and costs (including opportunity costs) for the affected groups, both by public authorities and household decision makers. Indeed, while health is the main preventive outcome for WASH, it is not a major motivating factor for WASH behaviours (Jenkins, 1999).

Sustaining impacts and achieving them at scale requires the continued wide acceptance and adoption of new technology, which may require additional promotional approaches. Sustainability and scalability of impacts are therefore central issues for policy and practice. Sustainability of impacts requires continued adherence by beneficiaries, solutions to “slippages” in behaviour and financial barriers to uptake, as well as technical solutions to ensure service delivery reliability. Scalability requires that impacts measured in small-scale efficacy settings (the “ideal settings” measured in many field trials) are achievable in the context of programme effectiveness (“real world” settings) where fidelity of implementation becomes crucial (Bamberger et al., 2010). For example, hygiene information, education and behaviour change activities are usually a component of most, if not all, programme designs which aim to scale-up service provision. However, there are concerns about whether these activities are being implemented in practice (Jimenez et al., 2014).

<sup>5</sup>Diarrhoeal disease is usually measured as the risk, incidence, or prevalence. Risk measures the probability of being ill during the measurement period. Incidence density or rate measures the average risk over the measurement period measured in average number of discrete disease spells. Longitudinal prevalence is more closely associated with duration of illness, usually measured as the proportion of days of illness during the measurement period. Longitudinal prevalence of diarrhoea is preferred on theoretical grounds and empirically is more strongly associated with child mortality and weight gain than incidence (Morris et al., 1996). Different technologies may also affect measures of incidence and prevalence differently. For example, Gross et al. (1989) noted that hygienic practices such as removal of faeces from the yard may have greater impact on spell duration than incidence.

However, the effectiveness of WASH technology in preventing disease transmission depends on both the biological efficacy of the technology and its acceptability and use, or effectiveness, among consumers in the environment where it is based (Eisenstein et al., 2007). Acceptability and use in turn are determined by the WASH promotional intervention, which motivates behaviour change by triggering drives (e.g., disgust), emotions (e.g., status) or interests (e.g., curiosity) (Biran et al., 2014; Aunger & Curtis, 2016). Authors of diarrhoea efficacy studies have referred to lack of convenience and limited observability of health benefits in explaining why compliance rates may be low for household water treatment (Quick et al., 2002). As far back as the 1960s, Rogers (2005) documented the low level of use of public spigots in Egypt, despite government media campaigns warning people of the risks from drinking canal water. Qualitative research suggested various causes, including that users did not like the chemical taste of the chlorinated water, rumours that the chemicals were being used to control fertility, women preferring to gather water from the canal banks where they socialised, and long queues, and fighting in the queues, due to low water pressure (Figure 4).

## 1.4 | Why this review is needed

### 1.4.1 | The policy debate and international targets

There is great interest in the impacts of WASH on child mortality in policy communities. This is in part due to the method of calculation of DALYs (Cairncross & Valdmanis, 2006), which sums years of life lost (YLL) and years lived with disability (YLD) associated with a particular exposure or disease. Every death attributed to infection, especially among children, is weighted heavily in YLL in the DALY calculation. In contrast, a calculation of YLD based on numbers of days experiencing diarrhoeal disease is rather smaller in endemic circumstances, since the typical child diarrhoeal risk among populations lacking access to clean drinking water may be three episodes per year (Clasen et al., 2015). For example, the recent GBD exercise estimates YLL for acute lower-respiratory tract infections at over 1,300 deaths per 100,000 and diarrhoea at 960 deaths per 100,000 (GBD, 2016 Cause of Death Collaborators, 2017a). These are the third and fourth highest numbers of YLL to a single disease among all

causes of mortality (and the highest among communicable diseases). In contrast, YLD were estimated at one-tenth of the level of YLLs for diarrhoea (100 per 100,000) and around 1% (10 per 100,000) of YLLs for lower-respiratory tract infections (GBD, 2017b).<sup>6</sup>

There has been broad consensus on the need for international targets to improve WASH technology access since the 1977 United Nations (UN) Water Conference at Mar del Plata and subsequent International Drinking Water Supply and Sanitation Decade of the 1980s (Jolly, 2004). The goal of that Decade, ratified by the Conference, was to provide adequate access to safe water and hygienic latrines to the population of the world by 1990 (Cairncross et al., 1980, p. xi). Yet, by 1990, only an estimated 76% of the global population were using an improved drinking water source and 54% used improved sanitation, as defined by the JMP (WHO/UNICEF, 2013). In 1990, the Convention on the Rights of the Child recognised the “right of the child to the enjoyment of the highest attainable standard of health... through the provision of... clean drinking water, taking into consideration the dangers and risks of environmental pollution” (Article 24, p. 57; cited in Jolly, 2004, p. 274).

The Millennium Declaration in 2000 included a water goal, and, following a declaration at the World Summit on Sustainable Development at Johannesburg in 2002, a sanitation goal was added (Jolly, 2004). The resulting Millennium Development Goal (MDG) 7 drinking water and sanitation targets were to halve (from 1990 levels) the proportion of people without sustainable access to safe drinking water and basic sanitation by 2015. The water indicator was later further defined as access to water from an improved source within 1 km of the household. This is roughly the time taken for a 30-min round-trip to collect water in the absence of queuing, which has been demonstrated as the time up to which basic needs for water supply can be reasonably met (Cairncross & Feachem, 2018; White et al., 1972). There are circumstances where it is likely that more than 30 min will be needed for 1 km roundtrips, such as mountainous or sandy terrain, or in water scarce regions where people may spend more time queuing at the water collection point than travelling to it (Dar & Khan, 2011).<sup>7</sup> It is worth noting that the apparatus has been in place to monitor progress on water collection times at national (rural and urban) in most countries at least since the Demographic and Health Surveys (DHSs) Phase II (1988–1993) included a question on the time taken to “go there, fetch water, and come back” (Institute for Resource Development/Macro International, 1990). JMP has since defined improved drinking water as “basic” when it requires <30 min round-trip to collect (Table 1).

The water target was declared met at the global level (WHO/UNICEF, 2013). However, by 2017, 144 million people used surface drinking water directly from a river, pond, canal or stream, 435

million people used unprotected wells, springs or other unimproved sources, and 206 million used improved water that required more than 30 min roundtrip to collect.<sup>8</sup> There also remain big regional inequalities in access. In sub-Saharan Africa, 416 million people use surface water, unimproved drinking water sources, or have limited access to improved services (requiring more than 30 min round-trip to collect). In South Asia, 137 million use surface water, unimproved water or have limited services, and in East Asia, 165 million people use them. The biggest improvements in access to drinking water have been in Asia, but coverage for 2.14 billion people in East Asia and the Pacific and 1.65 billion in South Asia remains “basic.” This means improved drinking water is provided at the community level or, if provided on premises, the supply is unreliable or contaminated.

The target for the MDG sanitation indicator, defined as the use of unshared, improved sanitation, was missed at the global level and in most countries in South Asia and sub-Saharan Africa by a wide margin (United Nations, 2015).<sup>9</sup> Of the 1.4 billion people who defecate in the open or use unimproved sanitation, 505 million live in South Asia (of which 375 million are in India) and 546 million in sub-Saharan Africa. A further 620 million share limited sanitation facilities with two or more households (233 million in South Asia, 188 million in sub-Saharan Africa and 145 million in East Asia and the Pacific). By 2015, 4.5 billion people lacked access to safely managed sanitation, where excreta were disposed of safely in situ or offsite (UN Water, 2018).

The Agenda for Sustainable Development set new global targets for 2030, enshrined in the Sustainable Development Goals (SDGs).<sup>10</sup> The SDGs include targets for both access to basic services, which is a necessary condition to improve quality of life outcomes, and use of improved drinking water and sanitation, which is the sufficient condition to improve them. The SDGs are more ambitious than the MDGs, aiming to “ensure the availability and sustainable management of water and sanitation for all” by 2030 (UN Water, 2018). This greater ambition is reflected in both the indicators being measured, going beyond “improved” to “safely managed” services (Table 1), and the targets, which in most cases require universality in coverage by 2030.<sup>11</sup> This greater ambition may be necessary to achieve the population health and nutrition improvements long claimed (Cumming et al., 2019). The SDGs also incorporated targets for hand washing for the first time.

Reaching these targets will be challenging, and not just for sanitation and hygiene. For example, only 15 countries with <95% coverage are on track to achieve universal coverage of basic drinking water, only 14 countries with <95% coverage are on track for universal basic sanitation, and only 18 countries are on track to eliminate open defaecation (WHO/UNICEF, 2017). In 2016, the UN proclaimed 2018–2028 the International Decade for Action on Water for Sustainable Development.<sup>12</sup>

<sup>6</sup>In addition, road injuries caused the fifth biggest numbers of YLL at 817 per 100,000 (of which pedestrian road injuries contributed 290 per 100,000) and were in the top 20 causes of YLD at around 200 per 100,000 (pedestrian injuries contributing one-quarter of these). While musculoskeletal disorders caused 31 YLL per 100,000, lower back and neck pain was the biggest single cause of YLD (over 1,000 per 100,000) and other musculoskeletal disorders were the seventh highest (over 500 YLD per 100,000). Animal contact was estimated to contribute 58 per 100,000 YLL and around 30 YLD per 100,000.

<sup>7</sup>A second issue with the water target, noted by Dar and Khan (2011), occurs where drinking water contaminated by chemicals may cause noninfectious diseases like arsenicosis or fluorosis.

<sup>8</sup>WASH access and use data from <https://washdata.org/>

<sup>9</sup>This relatively “uneven progress” in reaching WASH sector targets was in part due to the sanitation indicator, defined as unshared by households, being harder to reach than the water indicator, which included shared facilities at the community level (Cumming et al., 2014).

<sup>10</sup>See <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>

<sup>11</sup>Unlike other targets which specify 2030, the target for ODF was originally specified for 2025 (Hutton & Varughese, 2016).



To provide universal coverage, including appropriately serving the most disadvantaged people, it will be necessary to promote effective interventions for different groups, particularly disadvantaged groups who are most likely to be hidden from coverage, in the contexts in which they are used in private and public realms (e.g., schools, health facilities, places of transit, work, commerce and recreation, streets and fields).

It may appear difficult to understand the continued limited access to and use of WASH in spite of these commitments, when the technologies and resources exist to provide everyone with safely managed WASH, and improved WASH provides the foundation for combating communicable diseases like diarrhoea which is endemic in low-income communities, killing millions every year. Improved WASH is also important for blocking infectious disease transmission in epidemics, such as cholera outbreaks and the COVID-19 pandemic (Howard et al., 2020).

At least part of the reason is due to competing priorities among decision makers, whether they are policymakers at the top, service providers, or service users at the bottom. In order to stand a chance of meeting universal SDG targets, decision makers need access to evidence on what are the most effective ways to provide access to and promote use of WASH services, in particular contexts, and for specific groups.

## 1.4.2 | Existing systematic evidence

There has been an explosion in the production of studies like randomised controlled trials (RCTs) that are able to attribute changes in diarrhoeal disease to WASH interventions (Chirgwin et al., submitted). Correspondingly, many systematic reviews and meta-analyses have synthesised the effects of these studies in L&MICs. The earliest reviews covered faeces-related infections associated with water and sanitation provision including diarrhoea (Esrey et al., 1985, 1991). Esrey concluded that “safe excreta disposal and proper use of water for personal and domestic hygiene appear to be more important than drinking water quality in achieving broad health impacts” (Esrey et al., 1991, p. 31).

Fewtrell and Colford (2004), Fewtrell et al. (2005) updated Esrey et al., (1985, 1991), concluding that both hygiene education and water quality interventions reduced the risk of diarrhoea disease by about 40% each in L&MICs, while sanitation provision or water supply reduced the risk by only around 20% each. A meta-analysis conducted by Clasen et al. (2006, updated in 2015) also supported the finding that water treatment at POU, particularly filtration, was more effective in reducing diarrhoea risk than other types of water improvements. These findings were replicated in Hunter (2009) and the WHO (Wolf et al., 2014, 2018). Interventions appeared to be more effective when a safe water storage container was also provided (Clasen et al., 2015), as it is, for example, in filtration devices from which water is accessed through a tap. A few meta-analyses of higher quality studies also found that piped water to households

significantly reduced diarrhoea morbidity (Waddington et al., 2009; Wolf et al., 2018). Wolf et al. (2018) also defined piped water according to reliability and quality, finding big impacts, although only one study measured it.

The evidence on sanitation is mixed. First, until the last decade there were few impact evaluations of sanitation impact covering more than a small number of clusters. Second, previous reviews did not take clustering into account. Thus, earlier reviews estimated between 25% and 35% reductions in diarrhoea from sanitation (Fewtrell & Colford, 2004; Norman et al., 2010; Waddington et al., 2009; Wolf et al., 2018). Replacing on-site sanitation with water-based sewerage was estimated to reduce the incidence of diarrhoea by around 30%, though it may not always be a suitable solution given the maintenance costs (Norman et al., 2010). Meta-analyses suggested hand-hygiene interventions reduced reported diarrhoea morbidity by between 30 and 50% (Aiello et al., 2008; Cairncross et al., 2010; Curtis & Cairncross, 2003; Ejemot-Nwadiaro et al., 2015; Waddington et al., 2009; Wolf et al., 2018). Soap provision appeared to be particularly effective (Aiello et al., 2009; Waddington et al., 2009).

A common finding from meta-analysis of indirect study comparisons (that is, findings across different contexts) is that bundling WASH interventions together does not produce additional effects in comparison with single water, sanitation or hygiene interventions (Fewtrell & Colford, 2004). For example, the World Bank's Independent Evaluation Group (White and Gunnarsson, 2008, p. 17) concluded that “the health impact of combined methods has not been found to be stronger than any single approaches.”

However, there are concerns about the quality of evidence on the effectiveness of WASH interventions in reducing morbidity, due to concerns about self- and carer-reported infection, particularly where survey participants are exposed to repeated measurement in open (nonblinded) trials (Schmidt & Cairncross, 2009; Zwane et al., 2011). One advantage of water treatment technology with respect to conducting trials is that it is possible to blind participants—for example, by providing the plastic bottle but no instructions about storage for ultraviolet (UV) filtration (Conroy et al., 1996). Schmidt and Cairncross (2009) reported that blinded studies of household water treatment estimated impacts that were not significantly different from zero. Other reviews of household water treatment trials have found smaller or null effects once double-blinding was taken into account (Clasen et al., 2006, 2015; Hunter, 2009; Waddington et al., 2009; Wolf et al., 2018).

Others have noted that water treatment technologies were more effective where adherence was higher (Arnold & Colford, 2007; Clasen et al., 2015; Waddington et al., 2009). One review found that “water quality interventions conducted over longer periods tend to show smaller effectiveness, while compliance rates, and therefore impacts, appear to fall markedly over time” (Waddington et al., 2009; iii). Schmidt and Cairncross (2009) concluded that “widespread promotion of household water treatment is premature given the available evidence” (p. 986). There therefore has been, and still is, considerable controversy as to the role and

<sup>12</sup><https://www.unwater.org/new-decade-water/> (accessed November 17, 2020).

scalability of water treatment interventions in combating diarrhoeal disease.

Issues affecting the quality of self-reported diarrhoea morbidity may also affect hygiene interventions. Although no studies with double blinding of participants and outcome assessors have been conducted of hygiene interventions in L&MICs, blinding of outcome assessors is achievable, for example, where participants were provided children's reading material unrelated to hygiene (Luby et al., 2006). One systematic review found a smaller, but still statistically significant, 20% reduction in risk of diarrhoeal morbidity in blinded trials of hygiene (Ejemot-Nwadiaro et al., 2015).

It appears to be increasingly common to adjust for lack of blinding using Bayesian meta-analysis. Hunter (2009) proposed a bias correction procedure to water treatment studies drawing on coefficients from a meta-epidemiology study, presented in Wood et al. (2008). In the updated Cochrane drinking water treatment review by Clasen et al. (2015), similar bias correction factors were also applied, although the authors noted that "we urge caution in relying on these adjusted estimates since the basis for the adjustment is from clinical (mainly drug) studies that may not be transferable to field studies of environmental interventions" (p. 9). Wolf et al. (2018) also adjusted household water treatment and hygiene interventions for bias due to lack of blinding, but not water supply and sanitation, arguing that water supply and sanitation have recognised benefits over and above health impacts, whereas water treatment and hygiene "usually aim exclusively to improve health which is apparent to the recipient" (p. 512). It is worth noting that the correction factor for hygiene studies is particularly large, yielding a highly imprecise estimate (odds ratio [OR] = 0.90, 95% confidence interval [CI] = 0.37, 2.17; 33 studies) that is much bigger than the bias from single blinding estimated by Ejemot-Nwadiaro et al. (2015).

To summarise, a large number of systematic reviews and meta-analyses of impact evaluations have linked WASH to diarrhoeal disease. The common outcome indicator collected in health impact evaluations and systematic reviews is diarrhoea morbidity. Morbidity is presumably collected as a proxy for diarrhoea mortality, since it is easier to measure for financial and ethnical reasons (Briscoe et al., 1985). However, it may be a poor proxy for diarrhoea mortality. Diarrhoeal disease prevalence—number of days with diarrhoea over a period—is thought to be more closely correlated with mortality than diarrhoea incidence—number of distinct diarrhoea spells over a period (Morris et al., 1996; Schmidt et al., 2011). In addition, morbidity estimates may be affected by censoring of data, particularly in observational studies and cluster-RCTs where recruitment of individuals is done after randomisation, or in studies (including RCTs) where children of different ages, and therefore lengths of exposure, are followed-up concurrently.

An alternative approach is to evaluate impacts on mortality. Meta-epidemiological evidence suggests that bias due to self-reporting is not problematic for all-cause mortality, and to a lesser extent cause-specific mortality (Savović et al., 2012; Wood et al., 2008), especially if taken from vital registration systems rather than reported by "verbal autopsy" in carer surveys. However, mortality measurement is complicated in

prospective studies due to ethics and statistical power (Briscoe et al., 1985). It is unethical to let people die in the course of intervention research when oral rehydration salts (ORS) or medical treatment may be easily provided to severely ill children. Designing studies which can measure childhood mortality with statistical precision is also complicated as mortality is a sufficiently rare outcome outside of epidemics. It is possible to use the greater statistical power of meta-analysis to pool findings from studies in order to estimate statistically precise effects of rare outcomes (Waddington et al., 2018), which is the approach taken in this systematic review.

## 2 | OBJECTIVES

The objective of the systematic review is to answer two main review questions:

- (1) What are the effects of improved water, sanitation and hygiene access in L&MICs on:
  - all-cause mortality in childhood?
  - diarrhoea and infection-related mortality in childhood?
- (2) To what extent do the effects vary by contextual factors, such as geographic location and baseline environmental risk, factors relating to the study participants, such as age, sex and immunocompromised status and factors relating to the implementation of the study itself, including design, risk of bias assessment and length of follow-up?

## 3 | METHODS

### 3.1 | Criteria for considering studies

Table 2 summarises the criteria for inclusion of populations, interventions, comparators, outcomes, study designs, settings, language and time frame, which are discussed in this section.<sup>13</sup>

#### 3.1.1 | Types of studies

Eligible studies are impact evaluations, defined as programme evaluations or field experiments that used quantitative approaches applied to experimental or observational data to measure the average effect of participating in a WASH programme relative to a control or comparison group (counterfactual) representing what would have happened to the same group in the absence of the programme. Eligible impact evaluations may also test different intervention mechanisms or technologies (i.e., active controls).

<sup>13</sup>This protocol has also been registered with Prospero: "Water, sanitation and hygiene (WASH) for reducing mortality in childhood in low- and middle-income countries," CRD42020210694. Available at: [https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=210694](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=210694) (accessed December 3, 2020).

**TABLE 2** Summary of inclusion criteria

Criteria	Definition
Populations	Human populations in low- and middle-income countries, as defined by the World Bank at the time the research was carried out. Populations of any age, sex, gender, disability or socio-economic status were included. Populations in epidemics (e.g., cholera outbreak) were excluded
Interventions	Studies of WASH interventions and technologies (exposures) were eligible. Interventions included demand side (behaviour change communication, subsidies, microloans, legal measures), supply side (direct hardware provision, privatisation and nationalisation, small-scale independent provider involvement, improved operator performance), or combinations of demand and/or supply side (decentralisation). Eligible WASH technologies included water supply, water treatment and safe storage, sanitation, and/or hygiene
Comparators	Impact evaluations where the comparison/control group received no intervention (standard WASH access), a different WASH intervention, a double-blind placebo (e.g., nonfunctioning water filter), a single-blind (e.g., school textbooks), or a pipeline (wait-list)
Outcomes	Mortality in childhood: all-cause and cause-specific mortality due to diarrhoea and infection. Mortality may be measured as a primary study outcome, or harvested from prospective trial participant flow diagrams
Study design	Randomised controlled trials, prospective and retrospective nonrandomised studies, natural experiments and observational studies with control for confounding (e.g., case-controls)
Settings	WASH provided for use in the household, community, school, or health care facility
Language	Studies in English, French, Spanish and Portuguese. Studies in other languages were included where an English translation was available
Time frame	Studies published at any time; no study was excluded based on date of publication

The following study designs will be included:

- (1) Prospective quantitative evaluations where participants were assigned to intervention(s) at individual or cluster levels:
  - a. RCTs with randomised assignment of units at individual and household level (e.g., Han & Hlaing, 1989), or with cluster assignment at a higher level (village, township, school or health facility) (e.g., Clasen et al., 2015; Pickering et al., 2015), quasi-RCTs using quasi-randomised assignment of units (e.g., alternation of clusters listed alphabetically), and studies using randomised encouragement, providing promotional information about an intervention or technology that is universally available (e.g., Devoto et al., 2012).
  - b. Nonrandomised studies (NRSs) with assignment of units based on practitioner or participant selection and contemporaneous measurement of outcomes by investigators at pre- and posttest in treatment and comparison groups,<sup>14</sup> or contemporaneous measurement by investigators in treatment and comparison group at posttest only. These include prospective cohort studies (e.g., Ryder et al., 1985), studies using methods such as statistical matching (e.g., propensity score matching [PSM]) (e.g., Reese et al., 2019), or direct control for confounding in adjusted analysis (e.g., Cole et al., 2012). Cross-sectional studies that analysed the relationship between WASH technology interventions and outcomes, which compared self-selected participants within the same group, but did not use any methods to control for confounding (e.g., Gross et al., 1989) were excluded.
    - c. NRSs with measurement by investigators in treatment group at least six time points pre- and posttest (interrupted time-series) (Fretheim et al., 2015).
    - d. Cross-over trials where treatment and control or comparison are swapped (e.g., Kirchoff et al., 1985).
- (2) NRSs designed retrospectively—that is, after intervention had occurred—with selection on observables, including nonrandomised pipeline design, studies using cross-section data, and studies using panel data or pseudo-panels of repeated cross-sections with an intervention and comparison group, using methods to match individuals and groups statistically or control for observable confounding in adjusted analysis (e.g., Galiani et al., 2005).
- (3) Case-control designs, and other types of studies of WASH exposures, will be included (e.g., Hoque et al., 1999; Victora et al., 1988).
- (4) Natural experiments designed retrospectively with selection on unobservables:
  - a. Natural experiments using exogenous treatment assignment rules, including randomised natural experiments (with assignment by public lottery), and natural experiments where assignment was by random errors in implementation (e.g., Morris et al., 2004).
  - b. Regression discontinuity designs (RDDs) with prospective assignment to intervention and comparison groups based on a threshold on a continuous variable (e.g., number of cases of disease in a community, poverty index) (e.g., Spears, 2013; Ziegelhöfer, 2012) or a physical threshold such as an administrative boundary (Villar & Waddington, 2019).
  - c. Studies using multistage or multivariate approaches with identification of compliers based on exogenous variation (e.g., instrumental variables) or double-differences (e.g., Geruso & Spears, 2018).

<sup>14</sup>This designation also applies to RCTs with noncompliance that are analysed using ATET.

Study designs without a comparator receiving a different intervention or service (e.g., Israel, 2007) will be excluded, as will studies that did not control for confounding (e.g., Wagner & Lanoix, 1959; World Bank, 1998). Studies, or components of studies, that collected and analysed purely qualitative evidence will also be excluded.

### 3.1.2 | Types of participants

Eligible participants are children aged under 20 in a L&MIC, as defined by the World Bank at the time the research was carried out. This includes children of any age, sex, gender, disability, immunocompromised state, or socioeconomic status, provided the study was conducted in endemic conditions found regularly in L&MICs. Hence, studies that were conducted under outbreak conditions, such as cholera epidemics, will be excluded (e.g., Daniels et al., 1999; Snow, 1855).

### 3.1.3 | Types of interventions

Studies will be included that measure receipt of a clearly defined WASH intervention, or use of a WASH technology for household and personal consumption. Interventions will be excluded in food hygiene in the workplace such as a market (e.g., Sobel et al., 1998), methods to control faecal contamination by animals in the yard (e.g., Oberhelman et al., 2006), and vector control methods such as fly spraying (e.g., Chavasse et al., 1999; Emerson et al., 1999). Interventions primarily supporting farms or businesses such as dam construction (e.g., Duflo & Pande, 2007) will also be excluded, as will interventions for groundwater or irrigation management (e.g., Meenakshi et al., 2013). Likewise, flood and drought management interventions and river, lake, coastal zone and wetlands management will be omitted. Finally, cointerventions with a major non-WASH component will be excluded, such as those providing deworming chemotherapy (e.g., Miguel & Kremer, 2004) and nutrition interventions (e.g., Humphrey et al., 2019), although any WASH-only trial arms without co-interventions of such studies are eligible (e.g., Luby et al., 2018; Null et al., 2018).

### 3.1.4 | Types of settings

WASH provided for use in the household, community, school, or health care facility. Studies on medicalised hygiene (such as sterilising wounds) will be excluded.

### 3.1.5 | Types of outcome measures

The two primary outcomes are all-cause mortality and mortality due to infection including diarrhoeal disease, usually defined as three or more water stools in a 24-h period (Bacqui et al., 1991). All-cause mortality may be defined by carers in self-report, and/or clinicians. Mortality due to

diarrhoea and other infections may be defined by carers in verbal autopsy and/or clinicians, or collected from vital registries.

Outcomes data will be taken from two sources. The first is in studies that report childhood mortality as a primary outcome like case-control studies and those using DHS data (e.g., Charmarbagwala et al., 2004). However, as noted above, for prospective studies, mortality measurement is complicated ethics and statistical power (Briscoe et al., 1985). Mortality data are recoverable from prospective studies that report losses to follow-up (attrition) in sample populations due to mortality, per CONSORT standards (Moher, 1998; Moher et al., 2010). These studies will therefore form the second source of evidence on childhood mortality.

Eligible outcomes relate to a WASH intervention mechanism or exposure. For example, where some programme evaluations of CDD—an approach that is used to provide projects in multiple sectors such as infrastructure, education and health—do not give estimates of outcomes separately for WASH projects, these outcomes will be excluded.

### 3.1.6 | Publication language and date

In addition to English, studies published in French (Messou et al., 1997), Spanish (Instituto Apoyo, 2000) and Portuguese (e.g., Rasella, 2003) will be included. Studies published in other languages are eligible if an English translation is available. Studies published at any time are eligible, hence no study will be excluded based on date of publication.

## 3.2 | Search methods for identification of studies

The review is being done based on an evidence and gap map of WASH programmes in L&MICs (Chirgwin et al., submitted), for which searches and coding of outcomes were done originally in 2018, and updates performed in 2020.

### 3.2.1 | Electronic searches

Electronic searches include the following academic databases: CAB Abstracts, CAB Global Health, Cochrane Library, Econlit, Embase, ERIC, Ovid MEDLINE, Popline, Proquest Social Sciences Collection. In addition, searches include completed trials identified in the following trial registries (OpenTrials, WHO International Clinical Trials Registry Platform, 3ie's RIDIE, and AEA RCT registry). An example search string is given in Appendix A.

### 3.2.2 | Searching other resources

Organisational website and repository hand searches include: the Impact Evaluation Repository of the International Initiative for Impact Evaluation (3ie), the Asian Development Bank, African

Development Bank, Inter-American Development Bank, J-PAL evaluations, Innovations for Poverty Action, CEGA water and sanitation research projects, DFID Research for Development, IMPROVE International, IRC (WASH),<sup>15</sup> Oxfam, UNICEF, US Agency for International Development, WaterAid, and the World Bank Development Impact Evaluation (DIME) and IEG. Finally, the bibliographies of all included systematic reviews were checked to identify additional primary studies and systematic reviews.

Reference lists of books, reports and meta-evaluations to capture studies missed in electronic searches, particularly early studies, include: White et al. (1972), Saunders and Warford (1976), Feachem et al. (1978), Cairncross et al. (1980), WHO (1983), Khan et al. (1986), Briscoe et al. (1986), Charmarbagwala et al. (2004), White and Gunnarsson (2008) and Esteves Mills and Cumming (2016).

### 3.3 | Data collection and analysis

#### 3.3.1 | Selection of studies

Studies identified for selection are based on searches which were done at title and abstract by two authors working independently, using EPPI-reviewer's machine learning software (Thomas et al., 2010), as presented in Chirgwin et al. (submitted). Selection of studies at full text was done by two authors working independently (Chirgwin et al., submitted).

#### 3.3.2 | Data extraction and management

A standardised data extraction form will be used to collect descriptive data from all the included studies. This includes country, location (rural, urban, nationwide), participant age-group, WASH intervention and technology, study design, environmental contamination as represented by community water and sanitation access at baseline, risk of bias, effect size and standard error.

#### 3.3.3 | Assessment of risk of bias in included studies

Comprehensive critical appraisal will be done, including risk of bias and publication bias assessments (Dickersin, 1990). Study methods will be critically appraised using a risk of bias tool developed for this review (Appendix B), drawing on existing approaches (Eldridge et al., 2016; Higgins et al., 2016; Hombrados & Waddington, 2012; Sterne et al., 2016; Waddington et al., 2017; Jimenez et al., 2018). The following categories of bias will be assessed:

1. *Confounding*: baseline characteristics are similar in magnitude, unbalanced characteristics are controlled in adjusted analysis; for randomised approaches, adjustments to the randomisation were taken into account in the analysis (e.g., stratum fixed effects, pairwise matching variables); time-varying confounding such as differential adherence in sustained interventions.
2. *Selection bias into the study*: randomisation approach and allocation concealment for individual and cluster-randomisation. For NRSs, timing of follow-up.
3. *Attrition or selection bias out of study*: total attrition and differential attrition across study groups (presentation of average characteristics across treatments and comparisons, and reasons for losses to follow-up). In cluster designed studies, where respondents are not followed over time, assessment is needed of the sampling strategy.
4. *Departures from intended intervention due to performance bias*: no-shows and cross-overs, addressed using intention-to-treat (ITT) or the complier average causal effect; spillover effects addressed through geographical distance between treatment and control or comparisons; differential contamination by external programmes (treatment confounding) addressed through information about adherence behaviour.
5. *Departures from intended intervention due to motivation bias*: observational data versus experimental data with clear link to intervention (informed consent); repeated measurement (frequency and regularity of survey rounds); Hawthorne, John Henry effects and survey effects (Zwane et al., 2011).
6. *Errors in measurement of intervention and outcome*: length of recall, definition of intervention and outcome, timing of data collection (seasonality, or seasonal variation accounted for some other way), method of data collection (observed versus reported), blinding of outcome assessors and, where possible, participants.
7. *Biases in analysis and reporting*: pre-analysis plan or study protocol, reporting ITT alongside other estimators, blinding of data analysts.
8. *Unit of analysis error*: methods used to adjust standard errors to account for correlation of observations within clusters (e.g., cluster-robust standard errors).

It is important to recognise that risk of bias will refer to the likelihood of bias in the estimated mortality rate (MR), which may be collected from study participant flow, as opposed to the overall risk of bias in the study for the other (primary) outcomes of interest. MRs will be computed over a standard period, as mortality measurements will increase over longer exposure periods, all else equal.<sup>16</sup> For example, Gebre et al. (2011) and Siegel et al. (2004) used the following calculation for  $CMR_j$ , the crude MR in study  $j$  per 1,000 person-years at risk:

<sup>15</sup>IRC was originally known as the International Reference Centre for Community Water Supply until the mid-1980s when it changed its name to IRC International Water and Sanitation Centre, but as of 2014 simply goes by IRC. We use IRC (WASH) to distinguish it from the International Rescue Committee.

<sup>16</sup>This is particularly important for comparative measures of MRs (effect sizes) that are time sensitive, such as risk differences, but less important for ratio estimates. However, follow-up length will be collected from studies and included in meta-regression analysis as it has been shown to be correlated with effect sizes in a previous meta-analysis of diarrhoea morbidity (Waddington et al., 2009).



$$\text{CMR}_j = \frac{D_j}{\frac{t_j}{12} \left( N_j - \frac{D_j + M_j}{2} \right)} \times 1000, \quad (1)$$

where  $D_j$  is the number of deaths,  $t_j$  is the study follow-up period in months,  $N_j$  is the baseline sample size, and  $M_j$  is the number of people who permanently migrated out of the study area over the follow-up period. This will be applied to data collected from included studies. Age-specific MRs for children may also be calculated by replacing Equation (1) with the numbers of deaths and population shares among the specific age groups. Cause-specific MRs will be calculated by replacing  $D_j$  with numbers of deaths attributed to diarrhoea and/or infectious diseases, determined by recalled verbal autopsy or taken from vital registration data. An important issue affecting crude death rate calculations is that they are right-censored; that is, where data are collected contemporaneously among participants regardless of age, children born into the study and younger children have completed shorter durations than older children (e.g., White et al., 2005). This causes downwards bias in the estimate of mortality in any single trial arm, although the bias may be less problematic in randomised trials with contemporaneous data collection across arms. In these cases, the age-specific MR per 1000 live births may be calculated, which is not susceptible to censoring:

$$\text{MR}_j = \frac{D_j}{(B_j - B_j^D)} \times 1000, \quad (2)$$

where  $B_j$  is the number of live births and  $B_j^D$  the number of stillbirths.

### 3.3.4 | Measures of treatment effect

The main estimate of treatment used in this review will be the OR. OR is calculated from the two-by-two frequency table:

$$\text{OR} = \frac{p_t / (1 - p_t)}{p_c / (1 - p_c)}, \quad (3)$$

where  $p_t$  is the proportion in the treatment group and  $p_c$  the proportion in the comparison group. Where studies use regression methods, OR will be calculated as:

$$\text{OR} = \frac{(y_c + b) / (1 - (y_c + b))}{y_c / (1 - y_c)}, \quad (4)$$

which makes use of  $p_t = y_c + b$ , where  $y_c$  is the outcome mean in the control and  $b$  the regression coefficient on the treatment variable. In such circumstances, the standard error of the logarithm of OR is given by:

$$\text{se}(\ln \text{OR}) = \sqrt{\frac{1}{n_t(y_c + b)} + \frac{1}{n_t(1 - y_c - b)} + \frac{1}{n_c y_c} + \frac{1}{n_c(1 - y_c)}}. \quad (5)$$

Some studies report the risk ratio, RR:

$$\text{RR} = \frac{p_t n_t / n_t}{p_c n_c / n_c} = \frac{p_t}{p_c}, \quad (6)$$

with standard error of the natural logarithm of RR given by:

$$\text{se}(\ln \text{RR}) = \sqrt{\frac{1}{n_t p_t} - \frac{1}{n_t} + \frac{1}{n_c p_c} - \frac{1}{n_c}}, \quad (7)$$

where treatment and control risks are available, RR will be transformed into OR using:

$$\text{OR} = \text{RR} \frac{1 - p_c}{1 - p_t}, \quad (8)$$

where risks are not given, assumed risks,  $\hat{p}_t$  and  $\hat{p}_c$ , equal to the median treatment and control risks from any studies in the same country measuring that outcome, will be used<sup>17</sup>:

$$\text{OR} = \text{RR} \frac{1 - \hat{p}_c}{1 - \hat{p}_c \text{RR}}, \quad (9)$$

where the hazards ratio is used, it will be converted into RR using the following transformation (Shor et al., 2017):

$$\text{RR} = \frac{1 - e^{\text{HR} \ln(1 - p_c)}}{p_c}. \quad (10)$$

Inserting Equation (10) into (9), it can be shown that:

$$\text{OR} = \frac{1 - p_c + (p_c - 1)e^{\text{HR} \ln(1 - p_c)}}{-p_c e^{\text{HR} \ln(1 - p_c)}}, \quad (11)$$

where 95% CIs are reported instead of  $t$  or  $\text{se}(b)$ , the following will be used to calculate the standard error (Higgins & Green, 2011):

$$\text{se}(\text{OR}) = e^{\frac{\ln(\text{CI}^U) - \ln(\text{CI}^L)}{3.92}}, \quad (12)$$

where  $\text{CI}^L$  and  $\text{CI}^U$  are, respectively, the lower and upper limits of the 95% CI.

Where studies report independent treatment and control arms, data for mortality from each treatment-control comparison will be included. Where studies report multiple correlated effect sizes, for example, factorial studies comparing multiple treatment groups against a single control arm (e.g., Luby et al., 2018; Null et al., 2018), the control arms will be split by assuming the populations and deaths were evenly distributed between comparisons (affecting the precision of estimate, but not the effect size). This is to prevent studies with multiple results receiving greater weight than studies with only one effect estimate, or the inclusion of positively correlated effect sizes, which lead to underestimation of the summary variance (Borenstein et al., 2009a).

Where it is not possible to split control groups for multiple study arms, effect estimates may be combined into “synthetic effects,” by calculating an average effect, weighted by sample size, of the relevant pair-wise comparisons in these studies, and variance

<sup>17</sup>The formula to transform RR into OR used by Clasen et al. (2015), taken from Higgins et al. (2011), is:  $\text{RR} = \frac{\text{OR}}{1 - \hat{p}_c + \hat{p}_c \text{OR}}$ , where  $\hat{p}_c$  represents the estimated control risk.

accounting for the correlation between correlated comparison groups from the same study. The formula for the pooled variance is given as (Borenstein et al., 2009b):

$$\begin{aligned} \text{Var}\left(\frac{1}{N}\sum_{i=1}^N \ln(\text{OR})_i\right) &= \left(\frac{1}{N}\right)^2 \text{Var}\left(\sum_{i=1}^N \ln(\text{OR})_i\right) \\ &= \left(\frac{1}{N}\right)^2 \left(\sum_{i=1}^N se_i^2 + \sum_{i \neq j} r_{ij} \sqrt{se_i^2 se_j^2}\right), \end{aligned} \quad (13)$$

where  $N$  is the total number of effects, and  $r_{ij}$  is the correlation between effects, calculated as the mean of the correlation of treatment groups and the correlation of the control groups, and  $se_i$  the standard errors. The correlation between control arms is assumed equal to 1 where the same control group was used as comparator and 0 otherwise. The correlation between treatment arms is assumed to be 0 when combining results from different treatment groups and 1 when combining results from the same treatment groups over time. When combining results across different individuals in the same treatment group the correlation is assumed 0.5, which estimates variance at the mid-point between the two extreme cases of treating comparisons as independent (with correlation coefficient equal to 0) and most likely underestimating the variance, or treating them as perfectly correlated (correlation coefficient of 1) and most likely overestimating the variance (Waddington et al., 2009).

### 3.3.5 | Unit of analysis issues

Where study participants are grouped into correlated clusters of observations, the following error correction formula will be used to adjust standard errors (Higgins & Green, 2011; Waddington et al., 2012):

$$se(\ln(\text{OR}))' = se(\ln(\text{OR}))\sqrt{1 + (m - 1)\rho}, \quad (14)$$

where  $m$  is the average number of observations per cluster and  $\rho$  is the intra-cluster correlation coefficient and  $1 + (m - 1)\rho$  is the design effect (*Deff*). This adjustment will not be applied in clustered studies where outcomes of interest were defined at the cluster level (e.g., municipality MR).

### 3.3.6 | Dealing with missing data

Where deaths are not reported in any intervention arm, 0.5 will be added to all frequencies in order to calculate OR (Sanchez-Meca et al., 2003).

Usually, the intra-cluster correlation coefficient,  $\rho$ , is not reported. It will need to be imputed for studies not presenting cluster-adjusted standard errors, or, for example, where effect sizes are calculated from participant flow diagrams. In studies that calculate test statistics using cluster-robust standard errors, it may be possible to estimate the standard error using:

$$se(\ln(\text{OR}))' = \frac{\ln(\text{OR})}{t'}, \quad (15)$$

where  $t'$  is the test statistic for the effect size estimate (OR), calculated using cluster-robust methods. Where the study does not use cluster-robust methods, the value of  $\rho$  may be imputed using the following approach. The variance of OR, is calculated as:

$$V(\text{OR}) = se(\text{OR})^2. \quad (16)$$

Inserting Equation (16) into (14) and rearranging gives<sup>18</sup>:

$$\rho = \left(1 - \frac{V(\text{OR})'}{V(\text{OR})}\right) \frac{1}{m - 1}, \quad (17)$$

where  $V(\text{OR})'$  is calculated as the square of Equation (15) and  $V(\text{OR})$  the square of Equation (5).

### 3.3.7 | Assessment of heterogeneity

A measure of relative heterogeneity—the proportion of variance due to variation in the “true” effects over sampling variation, or  $I^2$  (Higgins & Thompson, 2002)—will be calculated (Borenstein et al., 2017):

$$I^2 = \frac{\tau^2/k}{s_{RE}^2} = \frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2}, \quad (18)$$

where  $\tau^2$  is the estimated between-study variance,  $k$  is the number of studies, and  $s_{RE}^2$  is the random effects average variance; under the assumption of equal study variance and sample size, this is equal to the within-study variance  $\frac{\sigma^2}{n}$  plus the estimated between-study variance variance  $\tau^2$  (Borenstein et al., 2009a).  $I^2$  is usually expressed as a percentage rather than a proportion. Absolute heterogeneity will be measured as the between-study variance using the method of DerSimonian and Laird (1986) (Borenstein et al., 2017):

$$\begin{aligned} \tau^2 &= \max\left\{0, \frac{Q - df}{\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}}\right\}, \text{ where } Q \\ &= \sum_{i=1}^k w_i (\hat{b}_i - \hat{\beta})^2 \sim \chi_{df=k-1}^2, \end{aligned} \quad (19)$$

where  $\tau^2$  is artificially constrained at zero if the value falls below zero (since a variance cannot be <0), and  $Q$  is the inverse-variance weighted sum of squares of the difference between effect sizes  $\hat{b}_i$  and their estimated mean  $\hat{\beta}$ .  $Q$  is a statistic that follows the  $\chi^2$  distribution with

<sup>18</sup>Schmidt et al. (2011) presented another way to calculate  $Deff = \frac{V(\text{OR})'}{V(\text{OR})}$ . However, this method does not allow adjustment by studies' known numbers of clusters and observations within clusters. Hence, where  $m$  is known, Equation (17) will be the preferred means of calculating  $\rho$ .

degrees of freedom  $df = k - 1$ , where  $Q$  represents the observed variation and  $df$  the expected variation based on sampling error alone.

### 3.3.8 | Assessment of reporting biases

Publication bias will be assessed using two methods. Direct tests for publication bias will be done in meta-regression accounting for whether the study was published in a peer-review journal, or another publication route such as a working paper or organisational report. Indirect testing of small study effects will use inspection of contour-enhanced funnel graphs (Peters et al., 2008) and formal regression tests (Egger et al., 1997). These tests are based on the assumption that there are weaker incentives for researchers and journals to publish smaller sample studies that do not show significant findings, because the cost of such studies is less and/or that authors of underpowered (small-sample) studies are more likely to undertake exploratory analysis (called “p-hacking”) in order to obtain publishable results.

### 3.3.9 | Data synthesis

The sample requirements to estimate effects on mortality with statistical precision would usually be beyond what is possible in these studies. The approach taken in this systematic review will therefore be to take advantage of the greater power afforded by statistical meta-analysis, to attempt to estimate precise pooled effects. Inverse variance weighted random effects meta-analysis will be used to synthesise the findings. A standard approach to meta-analysis will be followed, including sensitivity analysis by risk of bias, subgroup analysis by mortality causation.

The random effects pooled effect is calculated as the expected mean effect across the distribution of population effects, using a modified weighted average of the inverse of the variance incorporating two sources of sampling error—within-study and between-study variation. Each study weight is equal to the inverse of the within-study error variance of the individual study  $s_i^2/n_i$  plus the estimated between-study variance  $\tau^2$ . Since the weight for a single study is equal to the inverse of the sum of the within and between study variances, the expected variance of the random effects average  $s_{RE}^2$  is the inverse of the sum of the weights across the studies (Borenstein et al., 2009a):

$$s_{RE}^2 = \frac{1}{\sum_i^k \frac{1}{s_i^2/n_i + \tau^2}} \quad (20)$$

### 3.3.10 | Subgroup analysis and investigation of heterogeneity

Subgroups and moderator variables will be collected based on what is theoretically associated with mortality. Subgroups may include children of different age (Butz et al., 1984), sex and socioeconomic status, and mortality causation (all-cause versus

diarrhoea or other infectious diseases). Moderators will include WASH intervention technology, environmental risk as determined by water and sanitation availability to capture community threshold effects (Shuval et al., 1981), location, study design, risk of bias and length of follow-up (Waddington et al., 2009). Baseline water and sanitation will be determined by the type that was most frequently used in the control or comparison group. Following Fewtrell and Colford (2004), where the study does not report the baseline assessment, the value will be imputed for the relevant country, location and year from the Joint Monitoring Programme dataset.

Both bivariate moderator analysis and multivariate meta-regression will be used to investigate heterogeneity. For meta-regression, a general-to-specific approach will be used to determine the optimum meta-regression specification (Mukherjee et al., 1997).

### 3.3.11 | Sensitivity analysis

Sensitivity of the findings will be assessed for outliers and by stratifying meta-analysis for those studies which included mortality as a primary outcome, versus those reporting mortality in participant flow.

## ROLES AND RESPONSIBILITIES

- Content: Hugh Waddington and Sandy Cairncross are responsible for content.
- Systematic review methods: Hugh Waddington and Sandy Cairncross are responsible for systematic review methods.
- Statistical analysis: Hugh Waddington is responsible for statistical analysis.
- Information retrieval: Hugh Waddington and Sandy Cairncross are responsible for information retrieval, based on searches designed by John Evers, Information Retrieval Specialist with Campbell IDCG.

## SOURCES OF SUPPORT

The review is not supported financially, but is based on searches for an evidence and gap map that was funded by the Water Supply and Sanitation Collaborative Council (WSSCC).

## DECLARATIONS OF INTEREST

The authors have no vested interest in the outcomes of this review, nor any incentive to represent findings in a biased manner. Sandy Cairncross has been involved in the development of sanitation and hygiene interventions, Sandy Cairncross has also contributed to included impact evaluations, and Sandy Cairncross and Hugh Waddington have led previous systematic reviews and meta-analyses on WASH and diarrhoea.

## PLANS FOR UPDATING THE REVIEW

Hugh Waddington will be responsible for updating the review, which will be done once financial support is available.

## REFERENCES

- Abubakar, I., Aliyu, S. H., Arumugam, C., Usman, N. K., & Hunter, P. R. (2007). Treatment of cryptosporidiosis in immunocompromised individuals: Systematic review and meta-analysis. *British Journal of Clinical Pharmacology*, 63(4), 387–393. <https://doi.org/10.1111/j.1365-2125.2007.02873.x>
- Aiello, A. E., Coulborn, R. M., Perez, V., & Larson, E. L. (2008). Effect of hand hygiene on infectious disease risk in the community setting: A meta-analysis. *American Journal of Public Health*, 98(8), 1372–1381.
- Arnold, B. F., & Colford, J. M., Jr. (2007). Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhoea in developing countries: A systematic review and meta-analysis. *American Journal of Tropical Medicine and Hygiene*, 76(2), 354–364.
- Aunger, R., & Curtis, V. (2016). Behaviour centred design: Towards an applied science of behaviour change. *Health Psychology Review*, 10(4), 425–446. <https://doi.org/10.1080/17437199.2016.1219673>
- Bacqui, A. H., Black, R. E., Yunus, M. D., Hoque, A., A.R. Chowdhury, H. R., & Sack, R. B. (1991). Methodological issues in diarrhoeal diseases epidemiology: Definition of diarrhoeal episodes. *International Journal of Epidemiology*, 20(4), 1057–1063. <https://doi.org/10.1093/ije/20.4.1057>
- Bamberger, M., Rao, V., & Woolcock, M. (2010). *Using mixed methods in monitoring and evaluation: Experiences from international development* (Policy Research Working Paper 5245). Washington, D.C: The World Bank.
- Barreto, M., Genser, B., Strina, A., Marluca, Teixeira, M. G., Assis, A. M. O., Rego, R., Teles, C. A., Prado, M. S., Matos, S. M. A., Santos, D. N., dos Santos, L. A., & Cairncross, C. (2007). Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: Assessment by two cohort studies. *Lancet*, 370, 1622–1628.
- Benova, L., Cumming, O., & Campbell, O. M. (2014). Systematic review and meta-analysis: Association between water and sanitation environment and maternal mortality. *Tropical Medicine and International Health*, 19(4), 368–387.
- Bhatia, R., & Falkenmark, M. (1993). Water resources and the urban poor: Innovative approaches and policy imperatives. Working Paper 46877. The World Bank, Washington, D.C.
- Biran, A., Schmidt, W.-P., Varadharajan, K. S., Rajaraman, D., Kumar, R., Greenland, K., Gopalan, B., Aunger, R., & Curtis, V. (2014). Effect of a behaviour-change intervention on handwashing with soap in India (SuperAmma): A cluster-randomised trial. *Lancet. Globalization and health*, 2, e145–e154.
- Black, R., Cousens, S., Johnson, H. L., Lawn, J. E., Rudan, I., Bassani, D. G., Jha, P., Campbell, H., Walker, C. F., Cibulskis, R., Eisele, T., Liu, L., & Mathers, C. (2010). Global, regional, and national causes of child mortality in 2008: A systematic analysis. *The Lancet*, 375(9730), 1969–1987.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009a). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, 97–111. <https://doi.org/10.1002/jrsm.12>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009b). *Introduction to meta-analysis*. John Wiley and Sons.
- Borenstein, M., Higgins, J., Hedges, L. V., & Rothstein, H. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis. Methods*, 8(1), 5–18. <https://doi.org/10.1002/jrsm.1230>
- Bosch, C., Hommann, K., Rubio, G., Sadoff, C., & Travers, L. (2002). Water and sanitation. In J. Klugman (Ed.), *A sourcebook for poverty reduction strategies. Macroeconomic and sectoral approaches* (Vol. 2, pp. 371–404). Washington, DC: The World Bank.
- Briscoe, J., Feachem, R. G., & Rahaman, M. M. (1985). *Measuring the impact of water supply and sanitation facilities on diarrhoea morbidity: Prospects for case-control methods*. WHO/CWS/85.3. CDD/OPR/85.1. World Health Organization, Geneva.
- Briscoe, J., Feachem, R. G., & Rahaman, M. M. (1986). *Evaluating health impact: Water supply, sanitation, and hygiene education*. IDRC-248e. UNICEF, International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) and International Development Research Centre (IDRC), Canada.
- Butz, W. P., Habicht, J. P., & DaVanzo, J. (1984). Environmental factors in the relationship between breastfeeding and Infant mortality: The role of sanitation and water in Malaysia. *American Journal of Epidemiology*, 119(4), 516–525.
- Cairncross, S. (1992). Control of enteric pathogens in developing countries. In R. Mitchell (Ed.), *Environmental microbiology*, (157–189). John Wiley and Sons.
- Cairncross, S. (2004). *The case for marketing sanitation. Sanitation and Hygiene Series Field Note. Water and Sanitation Programme-Africa*. Washington, D.C: The World Bank.
- Cairncross, S., Carruthers, I., Curtis, D., Feachem, R., Bradley, D., & Baldwin, G. (1980). *Evaluation for Village Water Supply Planning*. Chichester: John Wiley & Sons.
- Cairncross, S., & Cliff, J. L. (1987). Water use and health in Mueda, Mozambique. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81, 51–54.
- Cairncross, S., Cumming, O., Schechtman, L., Velleman, Y., & Waddington, H. (2014). Health impacts of sanitation and hygiene. In P. Cross & Y Coombes (Eds.), *Sanitation and hygiene in Africa: Where do we stand?: Analysis from the AfricaSan Conference, Kigali, Rwanda*. IWA Publishing.
- Cairncross, S., & Feachem, R. G. (2018). *Environmental health engineering in the tropics: Water, sanitation and disease control* (3rd ed.). Routledge.
- Cairncross, S., & Valdmanis, V. (2006). Water supply, sanitation and hygiene promotion. In D. Jamison, J. Breman, A. Measham, G. Alleyne, M. Claeson, D. Evans, P. Jha, A. Mills & P. Musgrove (Eds.), *Disease control priorities in developing countries* (2nd ed.(771–792). Oxford University Press.
- Cairncross, S., Blumenthal, U., Kolsky, P., Moraes, L., & Tayeh, A. (1996). The public and domestic domains in the transmission of disease. *Tropical Medicine and International Health*, 1(1), 27–34.
- Cairncross, S., Hunt, C., Boisson, S., Bostoen, K., Curtis, V., Fung, I., & Schmidt, W.-P. (2010). Water, sanitation and hygiene for the prevention of diarrhoea. *International Journal of Epidemiology*, 39, i193–i205. <https://doi.org/10.1093/ije/dyq035>
- Cairncross, S., & Kinneer, J. (1992). Elasticity of demand for water in Khartoum, Sudan. *Social Science and Medicine*, 34(2), 183–189.
- Cameron, L., Shah, M., & Olivia, S. (2013). *Impact evaluation of a large-scale rural sanitation project in Indonesia* (Policy Research Working Paper 6360, Impact Evaluation Series No. 83). Washington, D.C: The World Bank.
- Campbell, O., Benova, L., Gon, G., Afsana, K., & Cumming, O. (2015). Getting the basic rights—The role of water, sanitation and hygiene in maternal and reproductive health: A conceptual framework. *Tropical Medicine and International Health*, 20(3), 252–267.
- Charmarbagwala, R., Ranger, M., Waddington, H., & White, H. (2004). *The determinants of child health and nutrition: A meta-analysis*. Mimeo. Washington, D.C: The World Bank.
- Chavasse, D. C., Shier, R. P., Murphy, O. A., Huttly, S. R., Cousens, S. N., & Akhtar, T. (1999). Impact of fly control on childhood diarrhoea in Pakistan: Community-randomised trial. *Lancet*, 353(9146), 22–25.
- Chirgwin, H., Cairncross, S., Zehra, D., & Waddington, H. (Submitted). The effectiveness of interventions promoting uptake of water, sanitation, and hygiene (WASH) in low- and middle-income countries: Evidence map. Campbell Evidence and Gap Map.
- Clasen, T. F., Alexander, K.T., Sinclair, D., Boisson, S., Peletz, R., Chang, H. H., Majorin, F., & Cairncross, S. (2015). Interventions to improve water quality



- for preventing diarrhoea. *Cochrane Database of Systematic Reviews*, 10, CD004794. <https://doi.org/10.1002/14651858.CD004794.pub3>
- Clasen, T. F., Bostoen, K., Schmidt, W.-P., Boisson, S., Fung, I. C. H., Jenkins, M. W., Scott, B., Sugden, S., & Cairncross, S. (2010). Interventions to improve the disposal of human excreta for preventing diarrhoea (review). *Cochrane Database of Systematic Reviews*, 2010(6), CD007180. <https://doi.org/10.1002/14651858.CD007180.pub2>
- Clasen, T. F., Roberts, I., Rabie, T., Schmidt, W.-P., & Cairncross, S. (2006). Interventions to improve water quality for preventing diarrhoea (review). *Cochrane Database of Systematic Reviews*, 2006(3), 1–90. <https://doi.org/10.1002/14651858.CD004794.pub2>
- Clasen, T., Boisson, S., Routray, P., Torondel, B., Bell, M., Cumming, O., Ensink, J., Freeman, M., Jenkins, Odagiri, M., Ray, S., Sinha, A., Suar, M., & Schmidt, W. -P. (2014). Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Glob Health*, 2014(2), e645–e653.
- Cole, E. C., Hawkey, M., Rubino, J. R., Crookston, B. T., McCue, K., Dixon, J., Maqelana, T., Cwayi, J., Adams, C., & Kim, J. (2012). Comprehensive family hygiene promotion in peri-urban Cape Town: Gastrointestinal and respiratory illness and skin infection reduction in children aged under 5. *South African Journal of Child Health*, 6(4), 109–117. <https://doi.org/10.7196/SAJCH.459>
- Cumming, O., Arnold, B. F., Ban, R., Clasen, T., Esteves Mills, J., Freeman, M. C., Gordon, B., Guiteras, R., Howard, G., Hunter, P. R., Johnston, R. B., Pickering, A. J., Prendergast, A. J., Prüss-Ustün, A., Rosenboom, J. W., Spears, D., Sundberg, S., Wolf, J., Null, C., ... Colford, J. M. (2019). The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: A consensus statement. *BMC Medicine*, 17, 173. <https://doi.org/10.1186/s12916-019-1410-x>
- Cumming, O., Elliott, M., Overbo, A., & Bartram, J. (2014). Does global progress on sanitation really lag behind water? An analysis of global progress on community- and household-level access to safe water and sanitation. *PLOS One*, 9(12), e114699. <https://doi.org/10.1371/journal.pone.0114699>
- Curtis, V., Kanki, B., Mertens, T., Traore, E., Diallo, I., Tall, F., & Cousens, S. (1995). Potties, pits and pipes: Explaining hygiene behaviour in Burkina Faso. *Social Science and Medicine*, 41(3), 383–393.
- Curtis, V., & Cairncross, S. (2003). Effect of washing hands with soap on diarrhoea risk in the community: A systematic review. *Lancet Infectious Diseases*, 3(5), 275–281.
- Dangour, A. D., Watson, L., Cumming, O., Boisson, S., Che, Y., Velleman, Y., Cavill, S., Allen, E., & Uauy, R. (2013). Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children (review). *Cochrane Database of Systematic Reviews*, 2013(8), CD009382. <https://doi.org/10.1002/14651858.CD009382.pub2>
- Daniels, N. A., Simons, S. L., Rodrigues, A., Gunnlaugsson, G., Forster, T. S., Wells, J. G., Hutwagner, L., Tauxe, R. V., & Mintz, E. D. (1999). First do no harm: Making oral rehydration solution safer in a cholera epidemic. *American Journal of Tropical Medicine and Hygiene*, 60, 1051–1055.
- Dar, O. A., & Khan, M. S. (2011). Millennium development goals and the water target: Details, definitions and debate. *Tropical Medicine and International Health*, 16(5), 540–544. <https://doi.org/10.1111/j.1365-3156.2011.02736.x>
- De Buck, E., Van Remoortel, H., Hannes, K., Govender, T., Naidoo, S., Avau, B., Vandeveegaete, A., Musekiwa, A., Vittoria, L., Cargo, M., Mosler, H.-J., Vandekerckhove, P., & Young, T. (2017). Approaches to promote handwashing and sanitation behaviour change in low- and middle-income countries: A mixed method systematic review. *Campbell Systematic Reviews*, 2017, 7–447. <https://doi.org/10.4073/csr.2017.7>
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.
- Devoto, F., Duflo, E., Dupas, P., Pariente, W., & Pons, V. (2012). Happiness on tap: Piped water adoption in urban Morocco. *American Economic Journal: Economic Policy*, 4(4), 68–99.
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *The Journal of the American Medical Association*, 263(10), 1385–1389. <https://doi.org/10.1001/jama.1990.03440100097014>
- Duflo, E., & Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2), 601–646.
- Duflo, E., Greenstone, M., Guiteras, R., & Clasen, T. (2015). *Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation* (Working Paper 21521). Cambridge, MA: National Bureau of Economic Research.
- Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 1997(315), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Eisenstein, E.L., Lobach, D.F., Montgomery, P., Kawamoto, K., & Anstrom, K.J. (2007). Evaluating implementation fidelity in health information technology interventions. AMIA 2007 Symposium Proceedings, pp. 211–215.
- Ejemot-Nwadiaro, R. I., Ehiri, J. E., Arikpo, D., Meremikwu, M. M., & Critchley, J. A. (2015). Hand washing promotion for preventing diarrhoea. *Cochrane Database of Systematic Reviews*, 2015(9), CD004265. <https://doi.org/10.1002/14651858.CD004265.pub3>
- Eldridge, S., Campbell, M., Campbell, M., Drahota, A., Giraudeau, B., Higgins, J., Reeves, B., & Siegfried, N. (2016). *Revised Cochrane risk of bias tool for randomized trials (RoB 2.0) Additional considerations for cluster-randomized trials*. <https://www.riskofbias.info/welcome/rob-2-0-tool/archive-rob-2-0-cluster-randomized-trials-2016>
- Emerson, P. M., Lindsay, S. W., Walraven, G. E., Faal, H., Bogh, C., Lowe, K., & Bailey, R. L. (1999). Effect of fly control on trachoma and diarrhoea. *Lancet*, 353(9162), 1401–1403.
- Esrey, S. A. (1987). *The effect of improved water supplies and sanitation on child growth and diarrheal rates in Lesotho*. A thesis presented to the Faculty of the Graduate School of Cornell University, Ithaca, NY.
- Esrey, S. A., Feachem, R. G., & Hughes, J. M. (1985). Interventions for the control of diarrhoeal diseases among young children: Improving water supplies and excreta disposal facilities. *Bulletin of the World Health Organization*, 63(4), 757–772.
- Esrey, S. A., Potash, J. B., Roberts, L., & Schiff, C. (1991). Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization*, 69(5), 609–621.
- Esteves Mills, J., & Cumming, O. (2016). *The impact of water, sanitation and hygiene on key health and social outcomes: Review of evidence*. New York: Department for International Development (DFID) Evidence paper. Sanitation and Hygiene Applied Research for Equity (SHARE) Consortium, London and UNICEF.
- Evans, W. D., Pattanayak, S. K., Young, S., Buszin, J., Rai, S., & Bihm, J. W. (2014). Social marketing of water and sanitation products: A systematic review of peer-reviewed literature. *Social Science & Medicine*, 110, 18–25.
- Feachem, R. G., Burns, E. & Cairncross, S. (Eds.), (1978). *Water health and development: An interdisciplinary evaluation*. Tri-Med Books.
- Fewtrell, L., & Colford, J. M. (2004). *Water, sanitation and hygiene: Interventions and diarrhoea: A systematic review and meta-analysis*. Washington, D.C: World Bank.
- Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., & Haller, L. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and metaanalysis. *Lancet Infectious Diseases*, 5, 42–52.
- Freeman, M. C., Garn, J. V., Sclar, G. D., Boisson, S., Medlicott, K., Alexander, K. T., Penakalapati, G., Anderson, D., Mahtani, A. G., Grimes, J. E. T., Rehfuess, E. A., & Clasen, T. F. (2017). The impact of sanitation on infectious disease and nutritional status: A systematic



- review and meta-analysis. *International Journal of Hygiene and Environmental Health*, 220, 928–949.
- Fretheim, A., Zhang, F., Ross-Degnan, D., Oxman, A. D., Cheyne, H., Foy, R., Goodacre, S., Herrin, J., Kerse, N., McKinlay, R. J., Wright, A., & Soumerai, S. B. (2015). A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. *Journal of Clinical Epidemiology*, 68, 324–333.
- Galiani, S., Gertler, P., & Scharrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality in Argentina. *Journal of Political Economy*, 113(1), 83–120.
- GBD 2016 Causes of Death Collaborators. (2017a). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, 390, 1151–1210.
- GBD 2016 Causes of Death Collaborators. (2017b). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, 390, 1211–1259.
- Gebre, T., Ayele, B., Zerihun, M., House, J. I., Stoller, N. E., Zhou, Z., Ray, K. J., Gaynor, B. D., Porco, T. C., Emerson, P. M., Lietman, T. M., & Keenan, J. D. (2011). Latrine promotion for trachoma: Assessment of mortality from a cluster-randomized trial in Ethiopia. *The American Journal of Tropical Medicine and Hygiene*, 85(3), 518–523.
- Geruso, M., & Spears, D. (2018). Neighborhood sanitation and infant mortality. *American Economic Journal: Applied Economics*, 10(2), 125–162.
- Graham, J. P., Hirai, M., & Kim, S.-S. (2016). An analysis of water collection labor among women and children in 24 Sub-Saharan African countries. *PLoS One*, 11(6), e0155981. <https://doi.org/10.1371/journal.pone.0155981>
- Gross, R., Schell, B., Molina, M. C., Leao, M. A., & Strack, U. (1989). The impact of improvement of water supply and sanitation facilities on diarrhea and intestinal parasites: A Brazilian experience with children in two low-income urban communities. *Revue saude publica*, 23(3), 214–220.
- Han, A. M., & Hlaing, T. (1989). Prevention of diarrhoea and dysentery by hand washing. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 83, 128–131.
- Havelaar, A., Boonyakarnkul, T., Cunliffe, D., Grabow, W., Sobsey, M., Giddings, M., Magara, Y., Ohanian, E., Toft, P., Chorus, I., Cotruvo, J., Howard, G., & Jackson, P. (2003). *Guidelines for drinking water quality water borne pathogens*. Geneva: World Health Organization.
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions, Version 5.0.0*. John Wiley and Sons.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558. <https://doi.org/10.1002/sim.1186>
- Higgins, J. P. T., Sterne, J. A. C., Savovi, J., Page, M. J., Hróbjartsson, A., Boutron, I., Reeves, B., & Eldridge, S. (2016). A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods. Cochrane Database of Systematic Reviews*, 10(Suppl 1), 29–31. <https://doi.org/10.1002/14651858.CD201601>
- Hombrados, J. G., & Waddington, H. (2012). *A tool to assess risk of bias for experiments and quasi-experiments in development research*, Mimeo. New Delhi: International Initiative for Impact Evaluation.
- Hoque, B. A., Chakraborty, J., Chowdhury, J. T. A., Chowdhury, U. K., Ali, M., El Arifeen, S., & Sack, R. B. (1999). Effects of environmental factors on child survival in Bangladesh: A case control study. *Public Health*, 113, 57–64.
- Howard, G., Bartram, J., Brocklehurst, C., Colford, J. M., Costa, F., Cunliffe, D., Dreifelbis, R., Spindel Eisenberg, J. N., Evans, B., Girones, R., Hrudefy, S., Willetts, J., & Wright, C. Y. (2020). COVID-19: Urgent actions, critical reflections and future relevance of 'WaSH': Lessons for the current and future pandemics. *Water and Health*, 18(5), 613–630.
- Humphrey, J. H. (2009). Child undernutrition, tropical enteropathy, toilets, and handwashing. *Lancet*, 374, 1032–1035.
- Humphrey, J. H., Mbuya, M. N. N., Ntozini, R., Moulton, L. H., Stoltzfus, R. J., Tavengwa, N. V., Mutasa, K., Majo, F., Mutasa, B., Mangwadu, G., Chasokela, C. M., Chigumira, A., Chasokwa, B., Smith, L. E., Tielsch, J. M., Jones, A. D., Manges, A. R., Maluccio, J. A., Prendergast, A. J., ... Nderecha, W. T. S. (2019). Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: A cluster-randomised trial. *Lancet Global Health*, 7, e132–e147.
- Hunter, P. R. (2009). Household water treatment in developing countries: Comparing different intervention types using meta-regression. *Environmental Science and Technology*, 43(23), 8991–8997.
- Hutton, G., Haller, L., & Bartram, J. (2007). Global cost-benefit analysis of water supply and sanitation interventions. *Journal of water and health*, 5(4), 481–502.
- Hutton, G., & Varughese, M. C. (2016). *The costs of meeting the 2030 Sustainable Development Goal targets on drinking water, sanitation, and hygiene*. Water and Sanitation Program Technical Paper, January 2016. The World Bank, Washington, D.C.
- Instituto Apoyo (2000). *Evaluacion de impacto y sostenibilidad de los proyectos de foncodes*.(1–192).
- Israel, D. (2007). Impact of increased access and price on household water use in urban Bolivia. *The Journal of Environment Development*, 16(58), 58–83.
- Jadhav, A., Weitzman, A., & Smith-Greenaway, E. (2016). Household sanitation facilities and women's risk of non-partner sexual violence in India. *BMC Public Health*, 16, 1139. <https://doi.org/10.1186/s12889-016-3797-z>
- Jenkins, M. W. (1999). *Sanitation promotion in developing countries: Why the latrines of Benin are few and far between*. PhD dissertation, 428 pages. UC Davis, CA.
- Jimenez, A., Cavill, S., & Cairncross, S. (2014). The neglect of hygiene promotion in developing countries, as shown by the global analysis and assessment of sanitation and drinking water survey. *Journal of Water, Sanitation and Hygiene for Development*, 4(2), 240–247.
- Jimenez, E., Waddington, H., Goel, N., Prost, A., Pullin, A., White, H., Lahiri, S., & Narain, A. (2018). Mixing and matching: Using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes. *Journal of Development Effectiveness*, 10(4), 400–421.
- Jolly, R. (2004). Clean water for all. In R. Black & H. White (Eds.), *Targeting development: Critical perspectives on the Millennium Development Goals*, (273–291). Routledge.
- Jones-Hughes, T., Peters, J., Whear, R., Cooper, C., Evans, H., Depledge, M., & Pearson, M. (2013). Are interventions to reduce the impact of arsenic contamination of groundwater on human health in developing countries effective? A systematic review. *Environmental Evidence*, 2(11), 1–32.
- Kar, K., & Chambers, R. (2008). *Handbook on community-led total sanitation*. Plan UK.
- Kawata, K. (1978). Water and other environmental interventions: The minimum investment concept. *The American Journal of Clinical Nutrition*, 31, 2114–2123.
- Khan, M. S. I., Matin, A., Hassan, M. M., & Qader, M. M. A. (1986). *Annotated bibliography on water, sanitation and diarrhoeal diseases: Roles and relationships* (Specialized Bibliography Series No. 12). Dhaka: International Centre for Diarrhoeal Disease Research.
- Kirchhoff, L. V., McClelland, K. E., Do Carmo Pinho, M., Araujo, J. G., De Sousa, M. A., & Guerrant, R. L. (1985). Feasibility and efficacy of in-home water chlorination in rural north eastern Brazil. *Journal of Hygiene*, 94(2), 173–180.

- Koolwal, G., & van deWalle, D. (2010). *Access to water, women's work, and child outcomes* (Policy Research Working Paper WPS5302). Washington, DC: The World Bank.
- Liu, L., Johnson, H. L., Cousens, S., Perin, J., Scott, S., Lawn, J. E., Rudan, I., Campbell, H., Cibulskis, R., Li, M., & Mathers, C., Black, R.E. on behalf of Child Health Epidemiology Reference Group of WHO and UNICEF. (2012). Global, regional, and national causes of child mortality: An updated systematic analysis for 2010 with time trends since 2000. *Lancet*, 379(9832), 2151–2161.
- Luby, S. P., Agboatwalla, M., Painter, J., Altaf, A., Billhimer, W., & Keswick, B. (2006). Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. *Tropical Medicine and International Health*, 11(4), 479–489.
- Luby, S. P., Rahman, M., Arnold, B. F., Unicomb, L., Ashraf, S., Winch, P. J., Stewart, C. P., Begum, F., Hussain, F., Benjamin-Chung, J., Leontsini, E., Naser, A. M., Parvez, S. M., Hubbard, A. E., Lin, A., Nizame, F. A., Jannat, K., Ercumen, A., Ram, P. K., ... Colford, J. M., Jr. (2018). Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: A cluster randomised controlled trial. *The Lancet Global Health*, 6, e302–e315.
- Majorin, F., Torondel, B., Chan, G. K. S., & Clasen, T. (2019). Interventions to improve disposal of child faeces for preventing diarrhoea and soil-transmitted helminth infection. *Cochrane Database of Systematic Reviews*, 2019(9), CD011055. <https://doi.org/10.1002/14651858.CD011055.pub2>
- Maslow, A. (1943). A theory of human motivation. *Psychological Review*, 50, 370–396.
- McSweeney, B. G. (1979). Collection and analysis of data on rural women's time use. *Studies in Family Planning*, 10(11/12), 379–383.
- Meenakshi, J. V., Banerji, A., Mukherji, A., & Gupta, A. (2013). *Impact of metering of agricultural tubewells on groundwater use and informal groundwater irrigation services markets in West Bengal, India*. Impact Evaluation Report 4. 3ie, New Delhi.
- Messou, E., Sangaré, S. V., Jossieran, R., Le Corre, C., & Guélain, J. (1997). Effect of hygiene and water sanitation and oral rehydration on diarrhoea and mortality of children under five in rural area of Cote d'Ivoire. *Bulletin de la Societe de Pathologie Exotique*, 90(1), 44–47.
- Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159–217.
- Moher, D. (1998). CONSORT: An evolving tool to help improve the quality of reports of randomized controlled trials. *Consolidated Standards of Reporting Trials*. *JAMA*, 279, 1489–1491.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340, 1–28. <https://doi.org/10.1136/bmj.c869>
- Morris, S. S., Cousens, S. N., Kirkwood, B. R., Arthur, P., & Ross, D. A. (1996). Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence? *American Journal of Epidemiology*, 144(6), 582–588.
- Morris, S., Olinto, P., Flores, R., Nilson, E., & Figueiró, A. (2004). Conditional cash transfers are associated with a small reduction in the rate of weight gain of preschool children in northeast Brazil. *Journal of Nutrition*, 134(9), 2336–2341.
- Mukherjee, C., White, H., & Wuyts, M. (1997). *Econometrics and data analysis for developing countries. Priorities for development economics*. Routledge.
- Norman, G., Pedley, S., & Takkouche, B. (2010). Effects of sewerage on diarrhoea and enteric infections: A systematic review and meta-analysis. *Lancet Infectious Diseases*, 10(8), 536–544. [https://doi.org/10.1016/S1473-3099\(10\)70123-7](https://doi.org/10.1016/S1473-3099(10)70123-7)
- Null, C., Stewart, C. P., Pickering, A. J., Dentz, H. N., Arnold, B. F., Arnold, C. D., Benjamin-Chung, J., Clasen, T., Dewey, K. G., Fernald, L. C. H., Hubbard, A. E., Kariger, P., Lin, A., Luby, S. P., Mertens, A., Njenga, S. M., Nyambane, G., Ram, P. K., & Colford, J. M., Jr. (2018). Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: A cluster-randomised controlled trial. *The Lancet Global Health*, 6, e316–e329.
- Oberhelman, R. A., Gilman, R. H., Sheen, P., Cordova, J., Zimic, M., Cabrera, L., Meza, R., & Perez, J. (2006). An intervention-control study of corralling of free-ranging chickens to control *Campylobacter* infections among children in a Peruvian periurban shantytown. *American Journal of Tropical Medicine and Hygiene*, 74(6), 1054–1059.
- Peters, J., Sutton, A., Jones, D., Abrams, K., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996.
- Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., & Alzua, M.-L. (2015). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial. *Lancet. Globalization and health*, 3, e701–e711.
- Poulos, C., Pattanayak, S. K., & Jones, K. (2006). *A guide to water and sanitation sector impact evaluations (English)*. Doing impact evaluations series no. 4. Washington, DC: The World Bank.
- Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M. C., Gordon, B., Hunter, P. R., Medlicott, K., & Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health*, 222, 765–777. <https://doi.org/10.1016/j.ijheh.2019.05.004>
- Quick, R., Kimura, A., Thevos, A., Tembo, M., Shamputa, I., Hutwagner, L., & Mintz, E. (2002). Diarrhea prevention through household-level water disinfection and safe storage in Zambia. *American Journal of Tropical Medicine and Hygiene*, 66(5), 584–589.
- Rabi, M., Alhassan, M. B., Ejere, H. O. D., & Evans, J. R. (2012). Environmental sanitary interventions for preventing active trachoma. *Cochrane Database of Systematic Reviews*, 2012(2), CD004003. <https://doi.org/10.1002/14651858.CD004003.pub4>
- Ramirez, A., Ranis, G., & Stewart, F. (1998). *Economic growth and human development*. (Queen Elizabeth House Working Paper Series Number 18). Oxford: University of Oxford.
- Rasella, D. (2003). Impacto do Programa Água para Todos (PAT) sobre a morbimortalidade por diarreia em crianças do Estado da Bahia, Brasil. *Cad. Saúde Pública*, Rio de Janeiro, 29(1), 40–50.
- Reese, H., Routray, P., Torondel, B., Sinharoy, S. S., Mishra, S., Chang, H. H., & Clasen, T. (2019). Assessing longer-term effectiveness of a combined household-level piped water and sanitation intervention on child diarrhoea, acute respiratory infection, soil-transmitted helminth infection and nutritional status: A matched cohort study in rural Odisha, India. *International Journal of Epidemiology*, 48(6), 1757–1767. <https://doi.org/10.1093/ije/dyz157>
- Rogers, E. (2005). *Diffusion of innovations* (5th ed.). The Free Press.
- Root, G. P. M. (2001). Sanitation, community environments, and childhood diarrhoea in rural Zimbabwe. *Journal of Health, Population, and Nutrition*, 19(2), 73–82.
- Ryder, R. W., Reeves, W. C., Singh, N., Hall, C. B., Kapikian, A. Z., Gomez, B., & Sack, R. B. (1985). The childhood health effects of an improved water supply system on a remote Panamanian island. *The American Journal of Tropical Medicine and Hygiene*, 34(5), 921–924.
- Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8(4), 448–467.
- Sansom, K., Franceys, R., Njiru, C., & Morales-Reyes, J. (2003). *Contracting Out Water and Sanitation Services - Vol. 2. Case studies and analyses of Service and Management Contracts in developing countries*, WEDC, Loughborough. Available at <https://wedc-knowledge.lboro>.

- ac.uk/resources/books/Contracting\_Out\_Water\_and\_Sanitation\_Services\_-\_Vol\_2\_-\_Complete.pdf. Accessed 31 August 2020
- Saunders, R. J., & Warford, J. J. (1976). *Village water supply: Economics and policy in the developing world*. A World Bank research publication. The John Hopkins University Press.
- Savović, J., Jones, H., Altman, D., Harris, R., Jüni, P., Pildal, J., Als-Nielsen, B., Balk, E., Glud, C., Glud, L., Ioannidis, J., Schulz, K., Beynon, R., Welton, N., Wood, L., Moher, D., Deeks, J., & Sterne, J. (2012). Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: Combined analysis of meta-epidemiological studies. *Health Technology Assessment*, 2012(35), 1–82.
- Schmidt, W.-P., & Cairncross, S. (2009). Household water treatment in poor populations: Is there enough evidence for scaling up now? *Environmental Science and Technology*, 43(4), 986–992.
- Schmidt, W.-P., Arnold, B. F., Boisson, S., Genser, B., Luby, S. P., Barreto, M. L., Clasen, T., & Cairncross, S. (2011). Epidemiological methods in diarrhoea studies—An update. *International Journal of Epidemiology*, 40, 1678–1692. <https://doi.org/10.1093/ije/dyr152>
- Shiffman, M. A., Schneider, R., Faigenblum, J. M., & Helms, R. (1978). Field studies on water sanitation and health education in relation to health status in Central America. *Prog. Wat. Tech.* 11(1/2), 143–150.
- Shor, E., Roelfs, D., & Vang, Z. M. (2017). The “Hispanic mortality paradox” revisited: Meta-analysis and meta-regression of life-course differentials in Latin American and Caribbean immigrants' mortality. *Social Science and Medicine*, 186, 20–33.
- Shuval, H. R., Tilden, R. L., Perry, B. H., & Grosse, R. N. (1981). Effect of investments in water supply and sanitation on health status: A threshold-saturation theory. *Bulletin of World Health Organization*, 59(2), 243–248.
- Siegel, J. S., Swanson, D. A., & Shryock, H. S. (2004). *The methods and materials of demography* (2nd ed.). Elsevier/Academic Press.
- Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). John Churchill.
- Sobel, J., Mahon, B., Mendoza, C. E., Passaro, D., Cano, F., Baier, K., Racioppi, F., Hutwagner, L., & Mintz, E. (1998). Reduction of fecal contamination of street-vended beverages in Guatemala by a simple system for water purification and storage, handwashing, and beverage storage. *American Journal of Tropical Medicine and Hygiene*, 59, 380–387.
- Sommer, M., Ferron, S., Cavill, S., & House, S. (2014). Violence, gender and WASH: Spurring action on a complex, underdocumented and sensitive topic. *Environment and Urbanization*, 27(1), 105–116. <https://doi.org/10.1177/0956247814564528>
- Sorenson, S. B., Morssink, C., & Campos, P. A. (2011). Safe access to safe water in low income countries: Water fetching in current times. *Social Science and Medicine*, 72, 1522–1526.
- Spears, D. (2013). *Essays in the economics of sanitation and human capital in developing countries*. PhD Thesis. Princeton, NJ.
- Stephenson, L. S., Latham, M. C., & Ottesen, E. A. (2000). Malnutrition and parasitic helminth infections. *Parasitology*, 121(Suppl.. 2000.S23–S38.
- Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., ... Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 2016(355), i4919.
- Thomas, J., Brunton, J., & Graziosi, S. (2010). *EPPI-Reviewer 4: Software for research synthesis*. EPPI-Centre Software. Social Science Research Unit, UCL Institute of Education, London.
- UN Water (2018). *Sustainable development goal 6 synthesis report on water and sanitation*. New York: United Nations.
- United Nations (2015). *The millennium development goals report 2015*. New York: United Nations.
- Victora, C. G., Smith, P. G., Vaughan, J. P., Nobre, L. C., Lombardi, C., Teixeira, A. M. B., Fuchs, S. C., Moreira, L. B., Gigante, L. P., & Barros, F. C. (1988). Water supply, sanitation and housing in relation to the risk of infant mortality from diarrhoea. *International Journal of Epidemiology*, 17(3), 651–654.
- Villar, P. F., & Waddington, H. (2019). Within-study comparison and risk of bias in international development: Systematic review and critical appraisal. Methods research paper. *Campbell Systematic Reviews*, 2019(15), e1027. <https://doi.org/10.1002/cl2.1027>
- Waddington, H., Aloe, A. M., Becker, B. J., Djimeu, E. W., Hombrados, J. G., Tugwell, P., Wells, G., & Reeves, B. (2017). Quasi-experimental study designs series-paper 6: Risk of bias assessment. *Journal of Clinical Epidemiology*, 89, 43–52. <https://doi.org/10.1016/j.jclinepi.2017.02.015>
- Waddington, H., Snilstveit, B., White, H., & Fewtrell, L. (2009). *Water, sanitation and hygiene interventions to combat childhood diarrhoea in developing countries. Synthetic review 1*. New Delhi: International Initiative for Impact Evaluation (3ie).
- Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., Bhavsar, A., Evers, J., Perez Koehlmoos, T., Petticrew, M., Valentine, J. C., & Tugwell, P. (2012). How to do a good systematic review of effects in international development: A toolkit. *Journal of Development Effectiveness*, 4(3), 359–387.
- Wagner, E. G., & Lanoix, J. N. (1958). *Excreta disposal for rural areas and small communities* (World Health Organization Monograph Series No. 39). Geneva: WHO.
- Wagner, E. G., & Lanoix, J. N. (1959). *Water supply for rural areas and small communities* (World Health Organization Monograph Series No. 42). Geneva: WHO.
- Waterkeyn, J., & Cairncross, S. (2005). Creating demand for sanitation and hygiene through community health clubs: A cost-effective intervention in two districts in Zimbabwe. *Social Science and Medicine*, 61, 1958–1970.
- White, G. R., Bradley, D. J., & White, A. U. (1972). *Drawers of water: Domestic water use in East Africa*. University of Chicago Press.
- White, H. (2004). Reducing infant and child death. In R. Black & H. White (Eds.), *Targeting development: Critical perspectives on the Millennium Development Goals* (pp. 1–26). Routledge.
- White, H., & Gunnarsson, V. (2008, June 30). What works in water supply and sanitation? Lessons from impact evaluations. A summary of findings. Washington, DC: World Bank. Independent Evaluation Group.
- White, H., Blöndal, N., Masset, E., & Waddington, H. (2005). *Maintaining momentum to 2015? An impact evaluation of interventions to improve maternal and child health and nutrition in Bangladesh*. Operations Evaluation Department. Washington, DC: World Bank.
- White, H., Menon, R., & Waddington, H. (2018). *Community-driven development: Does it build social cohesion or infrastructure? A mixed-method evidence synthesis*. 3ie Working Paper 30. 3ie, New Delhi.
- WHO/UNICEF. (2013). *Progress on sanitation and drinking-water: 2013 update*. Geneva: UNICEF and WHO.
- WHO/UNICEF. (2017). *Progress on drinking water, sanitation, and hygiene: 2017 update and SDG baselines*. Geneva, Switzerland: UNICEF and WHO.
- WHO/UNICEF. (2019). *Progress on drinking water, sanitation, and hygiene 2000-2017: Special focus on inequalities*. Geneva, Switzerland: WHO/UNICEF.
- WHO. (1983). *Minimum evaluation procedure*. Geneva: World Health Organization.
- Winter, S. C., & Barchi, F. (2016). Access to sanitation and violence against women: Evidence from Demographic Health Survey (DHS) data in Kenya. *International Journal of Environmental Health Research*, 26, 291–305. <https://doi.org/10.1080/09603123.2015.1111309>
- Wolf, J., Hunter, P. R., Freeman, M. C., Cumming, O., Clasen, T., Bartram, J., Higgins, J. P. T., Johnston, R., Medlicott, K., Boisson, S., & Prüss-Ustün, A. (2018). Impact of drinking water, sanitation and handwashing with soap on childhood diarrhoeal disease: Updated meta-analysis and meta-regression. *Tropical Medicine and International Health*, 23(5), 508–525.

- Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., Gluud, C., Martin, R. M., Wood, A. J. G., & Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *British Medical Journal*, 2008(336), 601–605. <https://doi.org/10.1136/bmj.39465.451748.AD>.
- World Bank. (1998). *Paraguay impact evaluation report: Community-based rural water systems and the development of village committees* (Report No. 17923). Operations Evaluation Department, The World Bank, Washington, DC.
- World Bank (2008). *Economic impacts of sanitation in Southeast Asia: A four-country study conducted in Cambodia, Indonesia, the Philippines and Vietnam under the Economics of Sanitation Initiative (ESI)*. *Water and Sanitation Program*. Washington, DC: World Bank.
- Wright, J., Gundry, S., & Conroy, R. (2004). Household drinking water in developing countries: A systematic review of microbiological contamination between source and point-of-use. *International Health*, 9(1), 106.
- Ziegelbauer, K., Speich, B., Mäusezahl, D., Bos, R., Keiser, J., & Utzinger, J. (2012). Effect of sanitation on soil-transmitted helminth infection: Systematic review and meta-analysis. *PLOS Medicine*, 9(1), e1001162. <https://doi.org/10.1371/journal.pmed.1001162>
- Ziegelhöfer, Z. (2012). *Down with diarrhea—Using fuzzy regression discontinuity design to link communal water supply with health* (IHEID Working Papers 05-2012). Economics Section, The Graduate Institute of International Studies.
- Zwane, A. P., Zinman, J., vanDusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D., Hornbeck, R., Gine, X., Duflo, E., Devoto, F., Crepon, B., & Banerjee, A. (2011). Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 108(5), 1821–1826.

**How to cite this article:** Waddington H, Cairncross S. PROTOCOL: Water, sanitation and hygiene for reducing childhood mortality in low- and middle-income countries. *Campbell Systematic Reviews*. 2021;17:e1135. <https://doi.org/10.1002/cl2.1135>

## APPENDIX A: EXAMPLE SEARCH STRING

The example search string is formatted for Ovid MEDLINE(R).

1. Developing Countries.sh,kf. (79987)
2. Africa/or Asia/or Caribbean/or West Indies/or South America/or Latin America/or Central America/(69677)
3. (Africa or Asia or Caribbean or West Indies or South America or Latin America or Central America).tw. (151966)
4. (Afghanistan or Albania or Algeria or Angola or Argentina or Armenia or Armenian or Azerbaijan or Bangladesh or Benin or Byelarus or Byelorussian or Belarus or Belorussian or Belorussia or Belize or Bhutan or Bolivia or Bosnia or Herzegovina or Hercegovina or Botswana or Brazil or Bulgaria or Burkina Faso or Burkina Fasso or Upper Volta or Burundi or Urundi or Cambodia or Khmer Republic or Kampuchea or Cameroon or Cameroons or Cameron or Camerons or Cape Verde or Central

African Republic or Chad or China or Colombia or Comoros or Comoro Islands or Comores or Mayotte or Congo or Zaire or Costa Rica or Cote d'Ivoire or Ivory Coast or Cuba or Djibouti or French Somaliland or Dominica or Dominican Republic or East Timor or East Timur or Timor Leste or Ecuador or Egypt or United Arab Republic or El Salvador or Eritrea or Ethiopia or Fiji or Gabon or Gabonese Republic or Gambia or Gaza or Georgia Republic or Georgian Republic or Ghana or Grenada or Guatemala or Guinea or Guiana or Guyana or Haiti or Honduras or India or Maldives or Indonesia or Iran or Iraq or Jamaica or Jordan or Kazakhstan or Kazakh or Kenya or Kiribati or Korea or Kosovo or Kyrgyzstan or Kirghizia or Kyrgyz Republic or Kirghiz or Kirgizstan or Lao PDR or Laos or Lebanon or Lesotho or Basutoland or Liberia or Libya or Macedonia or Madagascar or Malagasy Republic or Malaysia or Malaya or Malay or Sabah or Sarawak or Malawi or Mali or Marshall Islands or Mauritania or Mauritius or Agalega Islands or Mexico or Micronesia or Middle East or Moldova or Moldovia or Moldovan or Mongolia or Montenegro or Morocco or Ifni or Mozambique or Myanmar or Myanma or Burma or Namibia or Nepal or Netherlands Antilles or Nicaragua or Niger or Nigeria or Muscat or Pakistan or Palau or Palestine or Panama or Paraguay or Peru or Philippines or Philipines or Phillipines or Phillippines or Papua New Guinea or Romania or Rumania or Roumania or Rwanda or Ruanda or Saint Lucia or St Lucia or Saint Vincent or St Vincent or Grenadines or Samoa or Samoan Islands or Navigator Island or Navigator Islands or Sao Tome or Senegal or Serbia or Montenegro or Seychelles or Sierra Leone or Sri Lanka or Solomon Islands or Somalia or Sudan or Suriname or Surinam or Swaziland or South Africa or Syria or Tajikistan or Tadjikistan or Tadjikistan or Tadjhik or Tanzania or Thailand or Togo or Togolese Republic or Tonga or Tunisia or Turkey or Turkmenistan or Turkmen or Uganda or Ukraine or Uzbekistan or Uzbek or Vanuatu or New Hebrides or Venezuela or Vietnam or Viet Nam or West Bank or Yemen or Zambia or Zimbabwe).tw,sh. (1179705)

5. ((developing or less\* developed or under developed or under-developed or middle income or low\* income or underserved or under served or deprived or poor\*) adj (countr\* or nation? or population? or world or state\*).ti,ab. (78902)
6. ((developing or less\* developed or under developed or under-developed or middle income or low\* income) adj (economy or economies)).ti,ab. (400)
7. (low\* adj (gdp or gnp or gross domestic or gross national)).tw. (205)
8. (low adj3 middle adj3 countr\*).tw. (8913)
9. (lmic or Imics or third world or lami countr\*).tw. (5007)
10. transitional countr\*.tw. (139)
11. or/1-10 (1323187)
12. (sanitation or sewage or sewerage or wastewater or domestic water or (water adj2 (access\* or suppl\* or quality or quantit\* or standard\* or drinking)) or hygiene).ti,ab. (161588)
13. Sanitation/(6507)



14. water/or drinking water/or water supply/or water quality/or fresh water/(198561)
15. waste water/or sewage/(35209)
16. or/12-15 (343517)
17. hand hygiene/or hand disinfection/(5814)
18. (handwash\* or soap\* or (hand\* adj3 (wash\* or hygiene or clean\*)))ti,ab. (12639)
19. or/17-18 (14788)
20. respiratory tract infections/or whooping cough/(43029)
21. Sinusitis/(15930)
22. Common Cold/(4075)
23. Otitis Media/(16557)
24. Pharyngitis/(7508)
25. Influenza, Human/(43747)
26. laryngitis/or croup/(3881)
27. Epiglottitis/(947)
28. pneumonia/or bronchopneumonia/or pleuropneumonia/or pneumonia, bacterial/or pneumonia, pneumococcal/or pneumonia, staphylococcal/or pneumonia, viral/(66073)
29. bronchitis/or bronchiolitis/(22955)
30. (ARIs or (respiratory adj (disease\* or infection\* or illness\*)) or sinusitis or common cold\* or otitis media or pharyngitis or influenza or flu or coryza or laryngitis or epiglottitis or croup or pneumonia or bronchitis or bronchiolitis or pertussis or whooping cough).ti,ab. (304365)
31. or/20-30 (385850)
32. 11 and 19 and 31 (205)
33. (diarrh\* or dysenter\* or gastroenteritis or cholera\* or "waterborne infection\*" or enterotoxi\* or enteric or enteritis or "escherichia coli\*" or "e coli" or rotavirus\* or mortality or death\*).ti,ab. (1570626)
34. diarrhoea/or dysentery/or enteritis/or gastroenteritis/or cholera/or vibrio cholerae/or waterborne diseases/or enterotoxins/ or exp escherichia coli/or escherichia coli infections/or rotavirus/or exp mortality/or "causes of death"/or death/or fatal infections/(671938)
35. or/33-34 (1845583)
36. 11 and 16 and 35 (6244)
37. ((time adj3 (saving\* or allocat\* or consum\* or fetch\* or travel\*)) or (collect\* adj3 (time\* or behavio\* or chore\* or errand\* or drudgery or burden\* or inconvenien\*)))ti,ab. (65859)
38. Time Factors/(1109957)
39. 37 or 38 (1169839)
40. 11 and 16 and 39 (1713)
41. joint diseases/or arthritis/or exp musculoskeletal pain/or back pain/or low back pain/or neck pain/(97943)
42. ((back adj3 (pain\* or injur\*)) or backpain or ((neck or spine or spinal) adj3 (pain or injur\*)) or neckpain or musculoskeletal or (joint adj3 (pain or disease\* or injur\*)) or arthriti\* or osteoarthritis\*).ti,ab. (330866)
43. Amenorrhoea/or Menstruation/or Menstruation Disturbances/or Menstrual Cycle/or Oligomenorrhoea/or Dysmenorrhoea/(43803)
44. (menstruat\* or menses or menstrual or amenorrh\* or dysmenorrh\* or oligomenorrh\*).ti,ab. (58762)
45. 41 or 42 or 43 or 44 (453245)
46. 11 and 16 and 45 (255)
47. ("quasi experiment\*" or quasi-experiment\* or "random\* control\* trial\*" or "random\* trial\*" or RCT or (random\* adj3 allocat\*) or matching or "propensity score" or PSM or "regression discontinuity" or "discontinuous design" or RDD or "difference in difference\*" or difference-in-difference\* or "diff in diff" or DID or "case control" or cohort or "propensity weighted" or propensity-weighted or "interrupted time series" or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or "research synthesis" or "scoping review" or "rapid evidence assessment" or "systematic literature review" or "Systematic review" or "Meta-analy\*" or Metaanaly\* or "meta analy\*" or "Control\* evaluation" or "Control treatment" or "instrumental variable\*" or heckman or IV or ((quantitative or "comparison group\*" or counterfactual or "counter factual" or counter-factual or experiment\*) adj3 (design or study or analysis)) or QED).-ti,ab,kw. (3095201)
48. clinical trial/or clinical trial, phase i/or clinical trial, phase ii/or clinical trial, phase iii/or clinical trial, phase iv/or controlled clinical trial/or randomized controlled trial/or pragmatic clinical trial/(784504)
49. controlled clinical trials as topic/or nonrandomized controlled trials as topic/or randomized controlled trials as topic/or pragmatic clinical trials as topic/or case-control studies/or retrospective studies/or controlled before-after studies/or interrupted time series analysis/or random allocation/or cohort studies/or follow-up studies/or longitudinal studies/or prospective studies/or retrospective studies/or propensity score/ (2075906)
50. 47 or 48 or 49 (4824377)
51. 32 or 36 or 40 or 46 (8083)
52. 50 and 51 (1912)
53. limit 52 to yr="2014 -Current" (503)



## APPENDIX B: RISK OF BIAS TOOL

Question	Coding	Scoring criteria	Decision rules
<p><b>1. Confounding:</b> Was the allocation or identification mechanism able to address confounding</p>	Y, PY, PN, N, U		
<b>Confounding—justification</b>	Open answer	Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all subquestions, cite relevant pages)	
<ul style="list-style-type: none"> <li>• <b>Confounding: RCTs</b></li> </ul>	Y, PY, PN, N, U	<p><b>Are criteria adequately addressed?</b></p> <p><b>a)</b> The authors describe a random component in sequence generation/randomisation method (e.g., lottery, coin toss, random number table).*</p> <p><b>b)</b> if a special randomisation procedure is used to ensure balance, it is well described (stratification, pairwise matching, unique random draw, multiple random draws etc.) and adjustment is taken into account in the analysis (e.g., stratum fixed effects, pairwise matching variables).</p> <p><b>c)</b> if the unit of allocation was by beneficiary or group, there was some form of centralised allocation mechanism such as an on-site computer system to ensure adequate allocation concealment;</p> <p><b>d)</b> if a public lottery was used for the sequence generation, details were given on the exact settings and participants attending the lottery.</p> <p><b>e)</b> the unit of allocation is based on a sufficiently large sample size to equate groups on average, and</p> <p><b>f)</b> a balance table is reported for all subgroups receiving differential treatment within control or treatment groups, comparing means and standard deviations of variables; any unbalanced covariates at individual or cluster level are controlled in adjusted analysis, including cluster-level variables</p>	<ul style="list-style-type: none"> <li>- Score “Yes” if all criterion (a)–(f) are satisfied</li> <li>- Score “Probably Yes” if all criteria are satisfied but there is no balance table to justify the random allocation (f)</li> <li>- Score “Probably No” if (f) is not satisfied because there is no balance table reported and there is evidence suggesting a problem in the randomisation, such as baseline coefficients in a diff-in-diff regression table are very different or sample size is too small for the procedure used (using stratification when there are less than two units for each intervention and control group in each strata can lead to imbalance)</li> <li>- Score “No” if the sample size is not sufficient or any failure in the allocation mechanism could affect the randomisation process</li> <li>- Score “Unclear” if the paper does not provide details on the randomisation process, or uses a quasirandomisation process for which it is not clear has generated allocations equivalent to randomisation</li> </ul> <p>*If a quasi-randomised assignment approach is used (e.g., prospective assignment by alternation or alphabetical order), you must be sure that the process truly generates groupings equivalent to random assignment, to score “Yes” on this criteria. In order to assess the validity of the quasirandomisation process, the most important aspect is whether the assignment process might generate a correlation between participation status and other factors (e.g., gender, socioeconomic status, pre-existing health condition) determining outcomes; you may consider whether assignment is done at cluster level (centralised) and covariate balance is reported in determining this</p>
<ul style="list-style-type: none"> <li>• <b>Confounding: RDD</b></li> </ul>	Y, PY, PN, N, U	<p><b>Are criteria adequately addressed?</b></p> <p><b>a)</b> Information about the programme targeting criteria are known, presented in the paper, and used to justify the statistical approach. Demonstration of the relationship between the assignment variable (a continuous variable or a discrete scaled variable with sufficient points either side of the cut-off) and outcome is done using a graph of the assignment-outcome relationship. Appropriate functional form may include local linear regression at assignment threshold, or ordered polynomial. The treatment effect may be measured as a change in intercept and/or change in slope (regression-kink design)</p>	<ul style="list-style-type: none"> <li>- Score “Yes” if criteria (a)–(c) are all satisfied including all falsification tests in part (b)</li> <li>- Score “Probably Yes” if participants or practitioners are unblinded but there is low risk of them manipulating the assignment score, and falsification tests suggest no other problems</li> <li>- Score “Probably No” if there are important differences between individuals on both sides of the cut-off, and any confirmation or falsification tests suggest potential problems.</li> <li>- Score “No” if there are important differences between individuals on both sides of the cut-off, and all confirmation or falsification tests suggest potential problems</li> </ul>

- Score "Unclear" if confirmation and falsification tests are not reported

- b) The relationship between assignment variable and outcomes are unconfounded at the threshold. Support for this can be obtained by confirmation test of no discontinuity at the cut-off in terms of baseline characteristics around the threshold, and falsification tests such as:
- addition of a phase in which intervention is not present, or "placebo time period," for example, by estimating the pretest relationship between assignment variable and outcomes, as a falsification exercise. Responsiveness of the outcome variable to temporal changes in intervention can also help verify the functional form and to adjust for nonlinearities in the relationship
  - addition of a nonequivalent outcome, or "placebo outcome." That is, assessing the effect on a second outcome variable that the intervention should not influence, as a falsification exercise
  - use of "placebo discontinuity" tests showing no other discontinuities in the assignment variable within the bandwidth of interest, as a falsification exercise
1. Classification of intervention status is not affected by systematic manipulation of the assignment variable by participants or decision makers, as indicated by:
- assignment decision rule is concealed from participants and practitioners, or
  - assignment variable is nonmanipulable by participants, practitioners or other decision makers, or
  - assignment variable is measured with random error
- The study should report a histogram of the assignment variable to demonstrate that bunching does not occur around the threshold, and McCrary test should be reported

• **Confounding NRS using statistical matching:**

Open answer

**Are criteria adequately addressed?**

- a) Information about the programme targeting criteria are known, presented in the paper, and used to justify the statistical approach
- b) Matching is done on pretest (or time-invariant) characteristics, including the outcome measured at pretest; matches are geographically local; the variables used to match are relevant (e.g., demographic and socio-economic factors) to explain both participation and the outcome (so that there can be no evident differences across groups in variables that might explain outcomes); and, for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme\*
- c) With the exception of Kernel matching, the means of the individual covariates are equated for treatment and comparison groups after matching

- Score "Yes", if (a)-(c) (if relevant) are addressed.
  - Score "Probably yes" if the selection into the programme was done according to clear targeting rules, which are used as matching variables, but there are imbalances remaining after matching
  - Score "Probably no" if programme assignment was self-selected by participants and there is no baseline data available to match the participants or groups
  - Score "No" if matching was done based on variables that are likely to be affected by the programme or any other scenario that affects validity of (a)-(c)
  - Score "Unclear" if relevant variables are not included in the matching equation, or if matching is based on characteristics collected at endline; or if insufficient details are provided on cluster controls
- \* Accounting for and matching on all relevant characteristics is usually only feasible when the programme allocation rule is known and there are no errors of targeting. There are different ways in which covariates can be taken into account. Observable differences across groups can be taken into account as covariates in the framework of

a regression analysis (e.g., propensity-weighted least squares) or can be assessed by testing equality of means between groups. Differences in unobservable characteristics can be taken into account using DD, FE, or RE if the only characteristics which are unobserved are time-invariant

- Score "Yes" if (a)-(d) are addressed
- Score "Probably yes" selection into the program was done according to clear rules, and equal trends demonstrated, but baseline imbalances between groups remained
- Score "Probably no" if equal trends not satisfied, and programme allocation was due to participant self-selection
- Score "No" if equal trends not satisfied, programme allocation was self-selected by participants, and some relevant time-varying characteristics are not controlled
- Score "Unclear" if insufficient details are provided, for example, on testing the equal trends assumption; or if insufficient details are provided on cluster covariates

\*DD, FE, and RE regression models are sometimes complemented with matching strategies. This combination approach is superior since it only uses in the estimation the common support region of the sample size, reducing the likelihood of existence of time-varying unobservable differences across groups affecting outcome of interest and removing biases arising from time-invariant unobservable characteristics

\*\* Knowing allocation rules for the programme—or even whether the nonparticipants were individuals that refused to participate in the programme, as opposed to individuals that were not given the opportunity to participate in the programme—can help in the assessment of whether the covariates accounted for in the regression capture all the relevant characteristics that explain differences between treatment and comparison

- Score "Yes", if (a)-(d) (if relevant) are addressed
- Score "Probably yes" if one of the test required for criterion (b) is not reported but the other is, and the rest of the criterion are addressed and the exogeneity of the instrument is clear
- Score "Probably no" if exogeneity of the instrument is not convincing and appropriate tests are not reported
- Score "No" otherwise if any of the tests required for criterion (a)-(d) are reported and not satisfied
- Score "Unclear" if relevant confounders are controlled for but appropriate statistical tests are not reported; or if insufficient details are provided on cluster controls

• **Confounding: NRS using DD or FE or RE analysis of panel data:\***

Y, PY, PN,  
N, U

- Are criteria adequately addressed?**
- a) Outcomes are measured at pretest (before intervention) and posttest (after intervention) using the same approach
  - b) examination of secular trends in outcomes shows parallel trends across treatment and comparison groups during periods prior to intervention
  - c) the method is combined by well-conducted statistical matching done according to clear programme allocation rules (see above), and baseline imbalances, including in the outcome are shown to be small
  - d) a comprehensive set of individual time-varying characteristics is controlled, including any cluster-level covariates that may affect the impact of the programme (e.g., rainfall)\*\*

• **Confounding: instrumental variables estimation (IV)**

Y, PY, PN,  
N, U

- Are criteria adequately addressed?**
- a) an appropriate instrumental variable is used which is exogenously generated: for example, due to a "natural" experiment or random allocation. If the instrument is the random assignment of the treatment, or fuzzy discontinuity, the reviewer should also assess the randomisation procedure or discontinuity assignment, as above
  - b) The joint test for the instruments is significant at the level of  $F \geq 10$ , or if an  $F$  test is not reported, the authors report and assess whether the  $R^2$  (goodness of fit) of the participation equation is sufficient for appropriate identification; and the identifying instruments are individually significant ( $p \leq .01$ )
  - c) the study assesses qualitatively why the instrument only affects the outcome via participation (the exclusion restriction); where at least two instruments are used, the authors report on an over-identifying test ( $p \leq .05$  is required to reject the null hypothesis); and none of the covariate controls can be affected by participation

d) for cluster-assignment, authors control for external cluster-level factors that might confound the impact of the programme (e.g., weather, infrastructure, community fixed effects, and so forth) through multivariate analysis

<p><b>Selection bias:</b> was any differential selection into the study adequately resolved?</p>	<p>Y, PY, PN, N, U</p>	<p><b>Are criteria adequately addressed?</b>                  a) If the study design is prospective, follow ups are recorded for all eligible participant units from recruitment onwards (i.e., prior to treatment). This is best shown using a study flow diagram, or reporting sufficient information to construct one                  b) Where the unit of allocation in a prospective study was at group level (geographical/social/cluster unit), allocation was performed on all units at the start of the study, or participants and recruiters are blinded to allocation status                  c) In retrospective studies (designed after intervention has commenced) and prospective cluster studies where participants are chosen after cluster assignment has been conducted, if those recruiting participants (or those being recruited) are aware of assignment prior to their recruitment and this is unlikely to affect recruitment differentially (e.g., participants chosen randomly using a sampling frame based on census and response rate is high), and the evidence presented does not suggest differences in recruitment across arms or characteristics across individuals (based on balance tests)                  d) Where evidence suggests there is selection bias into the study due to censoring of data (e.g., immortal time bias), this is accounted for using appropriate statistical methods (e.g., propensity weighted regression, Heckman selection model, proportional hazards model)</p>	<ul style="list-style-type: none"> <li>- Score "Yes" if (a)-(d) (if relevant) are satisfied.</li> <li>- Score "Probably yes" if the study used prospective design with adequate concealment, but no (or an incomplete) study flow diagram is reported</li> <li>- Score "Probably no" if there are threats to adequate concealment, for example, individual participants were chosen after cluster assignment was conducted or known, and there are differences between characteristics of the two groups beyond those expected by chance alone</li> <li>- Score "No" if there is evidence of differential recruitment into study arms and differences in characteristics of groups not compatible with chance</li> <li>- Score "Unclear" if it is unclear whether participants could affect their assignment status in response to knowledge of an un concealed allocation mechanism, and no information is presented about participant characteristics</li> </ul>
--	----------------------------	--	--

<p><b>Selection bias—justification</b></p>	<p>Open answer</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all subquestions, cite relevant pages)</p>
<p><b>Attrition bias:</b> was any differential selection out of the study adequately resolved?</p>	<p>Y, PY, PN, N, U</p>	<p><b>Are criteria adequately addressed?</b>                  a) Total attrition (losses to follow up) between pre- and posttest in the study is &lt;20%                  b) Differential attrition is &lt;10 percentage points, and whether reasons for attrition are given and similar across groups                  c) Attrition is sufficiently low and similar reasons for attrition in treatment and control, or the study assesses losses to follow up to be random draws from the sample (e.g., by examining correlation with key characteristics across groups, or an F test of attrition on baseline characteristics and interacted with treatment status)                  d) Study participants are randomly sampled</p>

<p><b>Attrition bias—justification</b></p>	<p>Open answer</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p> <p><b>Motivation bias: was the process of being observed free from motivation bias?</b></p> <p>Y, PY, PN, N, U</p> <p><b>Are criteria adequately addressed?</b></p> <p>a) For data collected in the context of a particular intervention trial (randomised or nonrandomised assignment), the authors state explicitly that the process of monitoring the intervention and outcome measurement is blinded to participants and outcome assessors, or methods are used that would minimise risk of Hawthorne effects, John Henry effects or survey effects such as infrequent observation or outcome questionnaires not referring to the intervention. Authors may adapt also the study design to estimate possible survey and Hawthorne effects (e.g., a “pure control” with no monitoring except baseline end-line)</p> <p>b) Informed consent is not associated with a particular intervention, as in the case of a regular household survey or a cluster-RCT, data are collected from administrative records, or in the context of a retrospective (ex post) evaluation</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p> <ul style="list-style-type: none"> <li>- Score “Yes” if either criterion (a) or (b) are satisfied;</li> <li>- Score “Probably yes” if the study is based on data collected during a trial and there is no obvious issue with the monitoring processes (e.g., due to infrequent monitoring) but authors do not mention potential risks</li> <li>- Score “Probably no” if there was imbalance in the frequency of monitoring in intervention groups, which might have influenced behaviour in treatment and control differentially</li> <li>- Score “No” if authors do not use an appropriate method to prevent Hawthorne and John Henry effects (e.g., blinding of participants) or control them (infrequent measurement, methods to ensure consistent monitoring across groups, measurement using a “pure control”)</li> <li>- Score “Unclear” if it is not clear whether the authors use an appropriate method to prevent Hawthorne and John Henry effects (e.g., blinding of outcomes and, or enumerators, other methods to ensure consistent monitoring across groups)</li> </ul>
<p><b>Motivation bias—justification</b></p>	<p>Open answer</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p> <p><b>Deviations from intended interventions:</b></p> <p>Was the study adequately protected against spillovers, no-shows and cross-overs?</p> <p>Y, PY, PN, N, U</p> <p><b>Are criteria adequately addressed?</b></p> <p>a) There were no implementation issues that might have led the control participants to receive the treatment, or authors use intention-to-treat estimation</p> <p>b) The intervention is unlikely to spill-over to comparisons (e.g., participants and nonparticipants are geographically and/or socially separated from one another and general equilibrium effects are not likely), or the potential effects of spillovers were measured (e.g., variation in the % of units within a cluster receiving the treatment)</p> <p>c) There is no risk of substitution (differential contamination) by external programs (also called treatment confounding): participants are isolated from other interventions which might be received differentially between treatment and controls which could explain changes in outcomes</p> <p>d) Errors in implementation fidelity by the intervening body were not systematic, or unlikely to affect the outcome</p> <p>e) For continuous interventions, measurement is taken of adherence to treatment among participants alongside outcomes</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p> <ul style="list-style-type: none"> <li>- Score “Yes” if criterion (a)–(e) (if relevant) are satisfied;</li> <li>- Score “Probably yes” if there is no obvious problem but there is no information reported on potential risks related to spillovers or contamination in the control group or if there were issues with spillovers but they were controlled for or measured</li> <li>- Score “Probably no” if any of the criterion (a)–(e) are not satisfied but the scale of the issue is not clear</li> <li>- Score “No” if any of the criterion (a)–(e) (if relevant) are not satisfied and happened at a large scale in the study</li> <li>- Score “Unclear” if spillovers, no-shows, cross-overs, implementation fidelity, or adherence to continuous interventions, are not addressed clearly</li> </ul>
<p><b>Deviation from intended interventions—justification</b></p>	<p>Open answer</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p> <p><b>Measurement bias:</b></p> <p>Was the study free from biases in measurement</p> <p>Y, PY, PN, N, U</p> <p><b>Are criteria adequately addressed?</b></p> <p>a) The study is a prospective design or in a retrospective design, participation in the intervention is observed, or the intervention</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p> <ul style="list-style-type: none"> <li>- Score “Yes” if criterion (a)–(f) are satisfied;</li> <li>- Score “Probably yes” if there is a small risk related to any of (a), (b), (c), (d), (e) of (f) and there is no more information provided to justify</li> </ul>



<p>of intervention and outcomes?</p>	<p>clearly and consistently defined and misreporting by participants or enumerators is unlikely</p> <p><b>b)</b> Outcomes are clearly and consistently defined for all participants and outcome assessors in the study</p> <p><b>c)</b> Outcomes are measured through observation (rather than self-report), and outcome assessors are blinded to intervention or it is shown they are unbiased (e.g., spot-checks to validate)</p> <p><b>d)</b> For self-reported outcomes: respondents in the intervention group are not more likely to have accurate answers than controls due to recall bias</p> <p><b>e)</b> For self-reported outcomes: respondents do not have incentives to over/under report something related to their performance or actions, or researchers put in place mechanisms to reduce the risk of reporting bias (outcome assessors not involved in the implementation of the intervention and it is clear that their answers to the survey will not affect what they receive in the future), or authors have measured bias through falsification tests or measuring the effect on placebo outcomes in cases where there was a risk of reporting bias</p> <p><b>f)</b> Timing of the data collection did not differ between intervention and comparison group, the baseline data are not likely to be differentially affected by the time of intervention (e.g., due to seasonality)</p>	<p>the absence of bias, or if potential biases are measured, for example, with placebo outcomes, and found to be null</p> <p>Score "Probably no" if there clear risks related to (a), (b), (c), (d), (e) or (f)</p> <p>Score "No" if there clear risks related to (a), (b), (c), (d), (e), or (f) and it is clear that authors were not able to control for the bias</p> <p>Score "Unclear" if it there is a risk related to any of (a), (b), (c), (d), (e), or (f) and there is no more information provided to justify the absence of bias</p>
--------------------------------------	--	--

<p><b>Measurement bias—justification</b></p>	<p>Open answer</p>	<p>Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)</p>
<p><b>Analysis reporting bias: RCTs</b> Was the study free from selective analysis reporting?</p>	<p>Y, PY, PN, N, U</p>	<p><b>Are criteria adequately addressed?</b></p> <p><b>a)</b> A pre-analysis plan or trial protocol is published and referred to or the trial was preregistered or the outcomes were pre-registered</p> <p><b>b)</b> Authors report results corresponding to the outcomes announced in the method section (there is no outcome reporting bias)</p> <p><b>c)</b> Authors report results of unadjusted analysis and intention to treat (ITT) estimation, alongside any adjusted and treatment-on-the-treated/complier-average-causal-effects analysis)</p> <p><b>d)</b> Authors use the appropriate analysis method (use baseline data when available) and different treatment arms are differentiated in the analysis</p> <p><b>e)</b> Authors have reported all the analysis which could help understand the results and no other bias is assessed as unclear due to the lack of an important analysis (e.g., a balance table or a subgroup analysis)</p>
<p><b>Analysis reporting bias: NRS</b> Was the study free from selective analysis reporting?</p>	<p>Y, PY, PN, N, U</p>	<p><b>Are criteria adequately addressed?</b></p> <p><b>a)</b> a pre-analysis plan is published, especially for prospective NRS (but ideally also for retrospective studies)</p> <p><b>b)</b> authors use "common" methods of estimation (i.e., credible analysis method to deal with attribution given the data available);</p>

<p>c) There is no evidence that outcomes were selectively reported (e.g., results for all relevant outcomes in the methods section are reported in the results section);</p> <p>d) Requirements for specific methods of analysis:</p> <ul style="list-style-type: none"> <li>- For RDD, Researchers should analyse the change in slope and/or level using different band-widths around the threshold or functional form. The following should be prespecified as far as possible, and reported in sensitivity analysis: (a) selection of optimal bandwidth using existing data-driven routines; (b) selection of appropriate functional form for the relationship between assignment and outcome variables; and (c) robustness checks of other bandwidths and functional form specifications</li> <li>- For PSM and covariate matching: (a) Where over 10% of participants fail to be matched, sensitivity analysis is used to re-estimate results using different matching methods (Kernel Matching techniques); (b) For matching with replacement, no single observation in the control group is matched with a large number of observations in the treatment group, and authors take into account the use of control observations multiple times against the same treatment in the standard error calculation; (c) for PSM, Rosenbaum's test suggests the results are not sensitive to the existence of hidden bias; (d) different matching methods including varying sample sizes yield the same results</li> <li>- For IV models, the authors test and report the results of a Hausman test for exogeneity (<math>p \leq .05</math> is required to reject the null hypothesis of exogeneity)</li> <li>- For Heckman selection models, the coefficient of the selectivity correction term (Rho) is significantly different from zero (<math>p &lt; .05</math>)</li> </ul>		<ul style="list-style-type: none"> <li>- Score "Probably No" if (b) is addressed, but authors did not present results for all outcomes announced in the method section, or did not meet requirement (d) although reported</li> <li>- Score "No" if authors use uncommon or less rigorous estimation methods such as failure to conduct multivariate analysis for outcomes equations, or if some important outcomes are subsequently omitted from the results or the significance and magnitude of important outcomes was not assessed</li> </ul>
<p><b>Analysis reporting bias—justification</b></p>	Open answer	Justification for coding decision (include a brief summary of justification for rating, mentioning your response to all sub questions, cite relevant pages)
<p><b>Method used to address differences between UoA and UoR/UoT</b></p>	Open answer	Briefly describe methods used to adjust standard errors to account for correlation across units (e.g., cluster-robust standard errors reported). Unit of analysis (UoA) is the unit of observation (e.g., individual, household, community, village), unit of randomisation (UoR) is the unit of assignment to control or treatment groups (e.g., individual, household, community, village), and unit of treatment (UoT) is the level at which treatment happens (e.g., individual, household, community, village)
<p><b>Unit of analysis: RCTs</b> Is unit of analysis in cluster allocation addressed in standard error calculation?</p> <p><b>Unit of analysis: NRS</b> Are correlations between units addressed in standard error calculation?</p>	Y, PY, PN, N, U	<ul style="list-style-type: none"> <li>- Score "Not applicable" if it is there is no clustering in the design at household or group levels</li> <li>- Score "Yes" if UoA equals UoR, or UoA is not equal to UoR and standard errors are clustered at the UoR level, or data is collapsed to the UoR level</li> <li>- Score "Not reported/unclear" if not enough information is provided on the way the standard errors were calculated or what the unit of analysis is. In this case, authors should consider adjusting standard errors using variance inflation formula in sensitivity analysis</li> <li>- Score "No" otherwise. In this case, standard errors should be adjusted using variance inflation formula</li> </ul>
	Y, PY, PN, N, U	<ul style="list-style-type: none"> <li>- Score "Not applicable" if there is no clustering in the design at household or group levels.</li> <li>- Score "Yes" if UoA equals UoT, or if UoA is not equal to UoT and standard errors are clustered at the UoT level, or data are collapsed to the UoT level</li> <li>- Score "Not reported/unclear" if not enough information is provided on the way the standard errors were calculated or what the unit of analysis is</li> <li>- Score "No" otherwise. In this case, standard errors should be adjusted using variance inflation formula</li> </ul>