BMJ Open

Study design and rationale for the PAASIM project: a matched cohort study on urban water supply improvements and infant enteric pathogen infection, gut microbiome development and health in Mozambique

Karen Levy [1,2] Joshua V Garn,[3] Zaida Adriano Cumbe,[4] Bacelar Muneme,[4] Christine S Fagnant-Sperati [1] Sydney Hubbard,[2] Antonio Júnior,[4] João Luís Manuel,[5] Magalhães Mangamela,[6] Sandy McGunegill,[2] Molly K Miller-Petrie [1] Jedidiah S Snyder,[2] Courtney Victor,[2] Lance A Waller [7] Konstantinos T Konstantinidis,[8] Thomas F Clasen,[2] Joe Brown [9] Rassul Nalá,[10] Matthew C Freeman [2]

Correspondence to
Dr Karen Levy;
klevyx@uw.edu and
Dr Matthew C Freeman;
matthew.freeman@emory.edu

ABSTRACT

Introduction Despite clear linkages between provision of clean water and improvements in child health, limited information exists about the health impacts of large water infrastructure improvements in low-income settings. Billions of dollars are spent annually to improve urban water supply, and rigorous evaluation of these improvements, especially targeting informal settlements, is critical to guide policy and investment strategies. Objective measures of infection and exposure to pathogens, and measures of gut function, are needed to understand the effectiveness and impact of water supply improvements.

Methods and analysis In the PAASIM study, we examine the impact of water system improvements on acute and chronic health outcomes in children in a low-income urban area of Beira, Mozambique, comprising 62 sub-neighbourhoods and ~26 300 households. This prospective matched cohort study follows 548 mother–child dyads from late pregnancy through 12 months of age. Primary outcomes include measures of enteric pathogen infections, gut microbiome composition and source drinking water microbiological quality, measured at the child's 12-month visit. Additional outcomes include diarrhoea prevalence, child growth, previous enteric pathogen exposure, child mortality and various measures of water access and quality. Our analyses will compare (1) subjects living in sub-neighbourhoods with the improved water to those living in sub-neighbourhoods without these improvements; and (2) subjects with household water connections on their premises to those without such a connection. This study will provide critical information to understand how to optimise investments for improving child health, filling the information gap about the impact of piped water provision to low-income urban households, using novel gastrointestinal disease outcomes.

Ethics and dissemination This study was approved by the Emory University Institutional Review Board and the National Bio-Ethics Committee for Health in Mozambique. The pre-analysis plan is published on the Open Science Framework platform (https://osf.io/4rkn6/). Results will be shared with relevant stakeholders locally, and through publications.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This matched cohort study of an urban water supply improvement project will provide critical information about the health impacts of providing piped water and household connections to low-income households.

⇒ We employ rigorous measures of exposure and novel and objective outcome measures, including gut microbiome composition and molecular detection of enteric pathogens.

⇒ The study design allows for examination of both neighbourhood and household-level effects of water supply improvements.

⇒ As a natural experiment, we are unable to randomise the intervention, leading to potential residual confounding.

⇒ We are unable to examine the impacts of all aspects of the city-wide water improvement project, due to lack of comparable populations, and instead focus only on the low-income neighbourhoods.

INTRODUCTION

Large-scale provision of disinfected, treated drinking water is considered one of the greatest public health achievements of the 20th century[1] and played an important role in improving child health in high-income countries.[2] In low-income and middle-income countries (LMICs) with high burdens of infectious diseases, inadequate water, sanitation and hygiene (WASH) conditions are strongly associated with poor child health outcomes, including diarrhoeal diseases, which are responsible for >500 000 deaths of children less than 5 years of age annually.[3 4] Repeated enteropathogen infections, regardless of symptoms, can lead to stunting and linear growth faltering, a comorbidity that impacts 4.7% of the children in LMICs and is responsible for a 4.8-times increase in mortality.[5] In Mozambique, 27% of stunting is attributed to unimproved water and sanitation.[6]

Robust studies of the health impacts of community water supply are needed

Rapid urbanisation is occurring globally, with urban areas expected to account for 96% of the additional 1.4 billion human population by 2030[7] and 68% of the global population is expected to live in urban areas by 2050.[8] While sub-Saharan Africa is still predominantly rural, by 2050 the continent is projected to be 56% urban.[9] To cope with urban growth, expanded infrastructure and services in cities and peri-urban areas will be essential.[7] In 2017 alone, an estimated US$3.95 billion was distributed in official development assistance to WASH in Africa (47% to the Eastern and Southern African region, specifically), with the majority going to improve water supply and sanitation systems.[10] Implementation challenges in lower-income settings—such as intermittent service and pathogen intrusion in the distribution system due to pipe breaks, pressure drops or illegal connections—limit the potential for engineered systems to provide a continuous supply of treated drinking water directly to homes, in adequate quantities to improve hygiene.[11 12] Given the considerable investment in providing piped services to low-income communities, rigorous evaluation of community-scale water provision is critical, to understand the real-world effectiveness and health impact of such systems in low-income contexts.[13–15]

Despite the clear biological link between safe water and child health and development, limited information exists about the health impacts of large water infrastructure improvements in low-income settings. A small number of studies have evaluated upgrades from intermittent to continuous water delivery in urban areas,[16] or localised improvements to water quality at shared water points.[17] Other studies have evaluated sanitation interventions, without examining drinking water or combined water and sanitation interventions.[18] A recent review of interventions to improve water quality globally found no studies evaluating reliable piped-in water supplies delivered to households and specifically called for rigorous research to assess the health impact of reticulated water

supply systems.[19] The review concluded that 'there is currently insufficient evidence to know if source-based improvements such as protected wells, communal tap stands or chlorination/filtration of community sources consistently reduce diarrhoea'.([19] p2) A WHO review of drinking water and diarrhoeal disease in low-income and middle-income settings concurs, stating that 'Although evidence on the effects of providing drinking water of higher quality supplied on premises…on health is scarce, data on the effects of supplying safely managed WASH services in LMICs are completely missing'.([20] p56) One reason for this limited evidence is that community-scale interventions are difficult to study using randomised control trial (RCT) methodology—the gold standard for causal inference. It is often infeasible to randomise intervention groups due to policy, planning and engineering considerations, and lack of adequate comparison groups. As such, alternative quasi-experimental designs must be applied.[21–23]

The predominance of studies in the WASH sector have focused on household-level and compound-level interventions because they lend themselves more readily to RCT methodology. Outside of the few aforementioned studies of community-wide infrastructure improvements, evaluations of the health impact of water quality improvements in low-income settings often focus on household water treatment such as boiling, chlorinating or filtering water, with studies predominantly conducted in rural settings.[19] Results of these trials have been mixed, since household-based approaches have various limitations, including low uptake and inconsistent use,[24 25] post-treatment contamination[26–29] and a poor record of sustained use.[30 31] Household water treatment interventions do not increase water quantity and availability and typical household WASH interventions are likely insufficient to prevent growth faltering in most cases.[32–35]

It is crucial to assess the impact of community-scale infrastructure improvements,[22] as this is an area that is particularly relevant to inform local and national policy-makers, aid agencies and development banks.[36] The area of most rapid growth in water access is via piped water supply connections, not household water treatment, and larger infrastructure interventions are also critical to achieving the scale of water supply improvements necessary to make impactful changes.[37]

Objective measures of gut health are needed

A vast majority of the WASH studies use as their primary health outcome caregiver-reported diarrhoea, primarily because acute diarrhoeal illness is responsible for ~10–12% of all deaths in children less than 5 years of age.[38 39] However, self-reported diarrhoea is an unreliable outcome due to courtesy, social desirability and recall bias,[40] local definitions of diarrhoea,[41–45] other self-reporting issues,[40 46–48] and the multiple potential aetiologies of diarrhoea symptoms.[49] Such biases are especially problematic where interventions cannot be blinded as is mainly the case for water interventions. Shedding of

enteropathogens, organisms that cause acute gastrointestinal illness, provides an unambiguous indicator of current infection, and increasingly is being used in the WASH field.[32 50 51] Advances in diagnostic techniques make it feasible to test for a wide variety of enteric pathogens simultaneously.[52 53] It is useful to understand enteric pathogen infections because chronic and repeated enteric pathogen infections in the first 2 years of life—with or without symptomatic diarrhoea—are associated with serious morbidities, including gut impairment, growth shortfalls and cognitive deficits by ages 7–9 years.[54–60] Such outcomes can have profound impacts on the health, development and well-being of individuals, communities and entire countries.[61 62] Host-level gastrointestinal conditions affected by environmental determinants, such as gut microbiome composition, may also help explain the long-term sequelae of enteric infections. While there is evidence of differences in gut microbiome composition across different cultures, regions and populations,[63–68] and environmental conditions,[69] to date specific WASH determinants of these differences, such as access to piped water, have not been evaluated using explicit counterfactuals. Thus measures of gut microbial conditions may provide objective outcomes to more accurately measure the effect of WASH interventions[70 71] and capture long-term sequelae,[72] resulting in a more complete understanding of the health impacts.[73]

Overview of study

In the PAASIM study (*Pesquisa Sobre o Acesso à Água e a Saúde Infantil em Moçambique - Research on Access to Water and Children's Health in Mozambique*), we address a series of questions about the impact of community-level water system improvements on acute and chronic health outcomes in children in a low-income urban area of Mozambique. This matched-control cohort study follows mother–child dyads from late pregnancy through children 12 months of age, examining the impact of living in an area with an improved water network and/or having a household water connection on a variety of aspects of access to drinking water, microbes in the child gut (including both pathogens and other resident gut microbes) and ultimately downstream health outcomes (including diarrhoea prevalence and growth) (figure 1). This study will provide critical information for agencies who seek to understand how to optimise investments for improving child health, helping to fill the information gap about the impact of providing piped water to urban, low-income households by isolating the effects of major community-level water supply improvements on novel gastrointestinal disease outcomes.

METHODS AND ANALYSIS

Study setting and description of the intervention

Our study site is the coastal city of Beira, the second largest city in Mozambique (population ~530 000),[74] which serves as a gateway for both the central interior portion of the country and a trade corridor to neighbouring land-locked nations. The centre of Beira is bordered by unplanned, informal settlements inhabited by over 300 000 low-income residents.[75] A 2018 survey in Beira by Water & Sanitation for the Urban Poor (WSUP) showed that of 5643 respondents, only 28% had a household water connection, and among those households without a connection, 83% used their neighbour's tap as their main source of water (unpublished data, courtesy of WSUP). Therefore, improvement of water supply and delivery infrastructure is a priority.

The World Bank funded the Water Service & Institutional Support (WASIS-II) project in 2016 to address the low access to improved water supply in Mozambique,[76] investing US$140 million with the Mozambican public institutions FIPAG (responsible for the public and private investment programme in urban water supply systems that serves as the water utility in Beira) and AURA, IP (the water regulatory authority responsible for the economic regulation and consumer protection of service provision). In addition, improvements in Beira are being augmented by investments from other groups, in particular the Dutch government, through infrastructure upgrades as well as emergency response funds following Cyclone Idai in 2019.[77] Improvements in the city of Beira include rehabilitation of water treatment facilities, replacing existing pipe mains that are failing, reticulation of water supply to new areas previously without water service, improving service in areas with poor coverage or low water pressure and subsidising water connection fees for the poor.

Study design

Several aspects of a city-wide water supply improvement project pose challenges to implementing a rigorous epidemiological study, particularly as people living in neighbourhoods with piped water often differ in myriad ways from people living in neighbourhoods without piped water. Water improvements to communities are often based on the needs or demographics of the community, government or donor priorities and engineering considerations. Provisions of water supply to new areas previously without water service—or dramatic improvements in access and availability—represent a fundamental development that changes the livability and sometimes the makeup of the community.[2] These issues lead to difficulty in finding a comparable control group

| Drinking water supply | Drinking water exposure | Gastrointestinal conditions | Health impact |
| Access to improved network and/or household water connection | Drinking water that is accessible, on premises, available when needed, and free from contamination | Improved acute and chronic child gastrointestinal conditions | Improved child growth and development |

Figure 1 PAASIM theory of change.

for epidemiological comparison. Furthermore, rollouts of water interventions often happen in continuous phases over time, and these changes might coincide with other events or community improvements that similarly impact health, making it difficult for that community to serve as its own control in a pre–post design.

By using a prospective matched cohort in this unique context of an ongoing natural experiment, we are able to overcome many of these difficulties. The prospective nature of our study allows better control of confounding through matching, restriction and rigorous and thoughtful collection of potential control variables. We specifically focus on one region of the city (figure 2), where some neighbourhoods received water system improvements focused on preventing water losses through replacement of the piped water distribution system in dense, low-income settings; other neighbouring areas with similar demographic characteristics did not receive these improvements. Neighbourhood-level matching took place in the context of a natural experiment, where the delayed rollout across the city allows us to find and compare intervention and control neighbourhoods that are similar in many ways, before the rollout eventually reaches all potential control areas.

The Water Loss Reduction Project represents a subset of the improvements being carried out by FIPAG, with co-funding from the Dutch government and the World Bank WASIS-II project. The improvements are designed to reduce illegal connections, thereby increasing the water pressure and quality and increasing the system's capacity for household connections. These areas also received some benefits related to improvements to the water intake and distribution systems. FIPAG undertook a campaign to offer new connections to households. These improvements were completed in some informal settlements in these low-income areas of the city in 2019, with other adjacent neighbourhoods with similar density and socioeconomic profile slated for completion in future years but not within the time frame of our study. These specific distribution system upgrades therefore represent a unique opportunity to examine the impacts of community-scale water improvements with neighbouring communities who did not receive the intervention serving as control areas for comparison. The neighbourhoods under study are also in the lowest income—and therefore highest need areas of the city.

We will perform analyses that take into account the four factorial possible household types, based on mother–child dyad living in a sub-neighbourhood with or without the improved water network and with or without a household connection (figure 3). Our primary analyses focus on assessing (A) the *total network effect*, by comparing subjects living in sub-neighbourhoods with the improved water network (Household Types 1 and 2) to those living in sub-neighbourhoods without these improvements (Household Types 3 and 4); and (B) the *direct household connection effect*, by comparing subjects with household water connections on their premises (Household Types 1

and 3) to those without a connection (Household Types 2 and 4). Depending on the results of these two primary analyses, secondary analyses may evaluate the other comparisons depicted in figure 3.

The reasons for some neighbourhoods receiving the improvements and others not were a result of resource constraints. According to FIPAG, decisions on the order of improvements in different neighbourhoods were guided by resource constraints as well as engineering logistics. We conducted a population-based survey (described below) that allowed us to both restrict and match study sub-neighbourhoods, thereby creating a statistically appropriate counterfactual for strong internal validity. The evaluation of a real-world intervention delivered in an informal urban setting provides strong external validity for estimating the effects of similar interventions in other LMIC urban sites. Our study design allows us to isolate the effects of both overall water supply infrastructural improvements as well as the presence of a household water connection. The presence of control areas not receiving upgrades adjacent to intervention areas that are matched on socioeconomic and density variables is unique to this study location. We collect data at multiple time points for each study household, allowing us to examine variability in each of the measures taken from each household, rather than at a single point in time, and also allowing for longitudinal analyses of the households and the individual enrolled subjects. We also employ rigorous measures of exposure and novel and objective outcome measures, including gut microbiome composition and molecular detection of enteric pathogens.

Patient and public involvement

The executive secretary of AURA, IP, was directly involved in the formation of the research questions, and FIPAG personnel were also engaged from the initiation of the project in helping to develop the study design. Our team also received input from other public agency stakeholders during workshops that were held prior to initiation of the study. Study subjects and members of the general public were not involved in the study design. We provide regular updates with data summaries to public agency stakeholders, and plan to disseminate the main results to all study participants and also through public presentations for stakeholders in both Beira and Maputo.

Sub-neighbourhood selection

Sub-neighbourhood eligibility, selection and matching of intervention and control sub-neighbourhoods occurred through a two-step process:

Intervention designation

Our study is a natural experiment, where the investigators had no control over the selection or timing of the intervention implementation. The study flow diagram is shown in figure 4. We worked with FIPAG to determine which neighbourhoods in Beira were to receive water distribution system upgrades prior to initiation
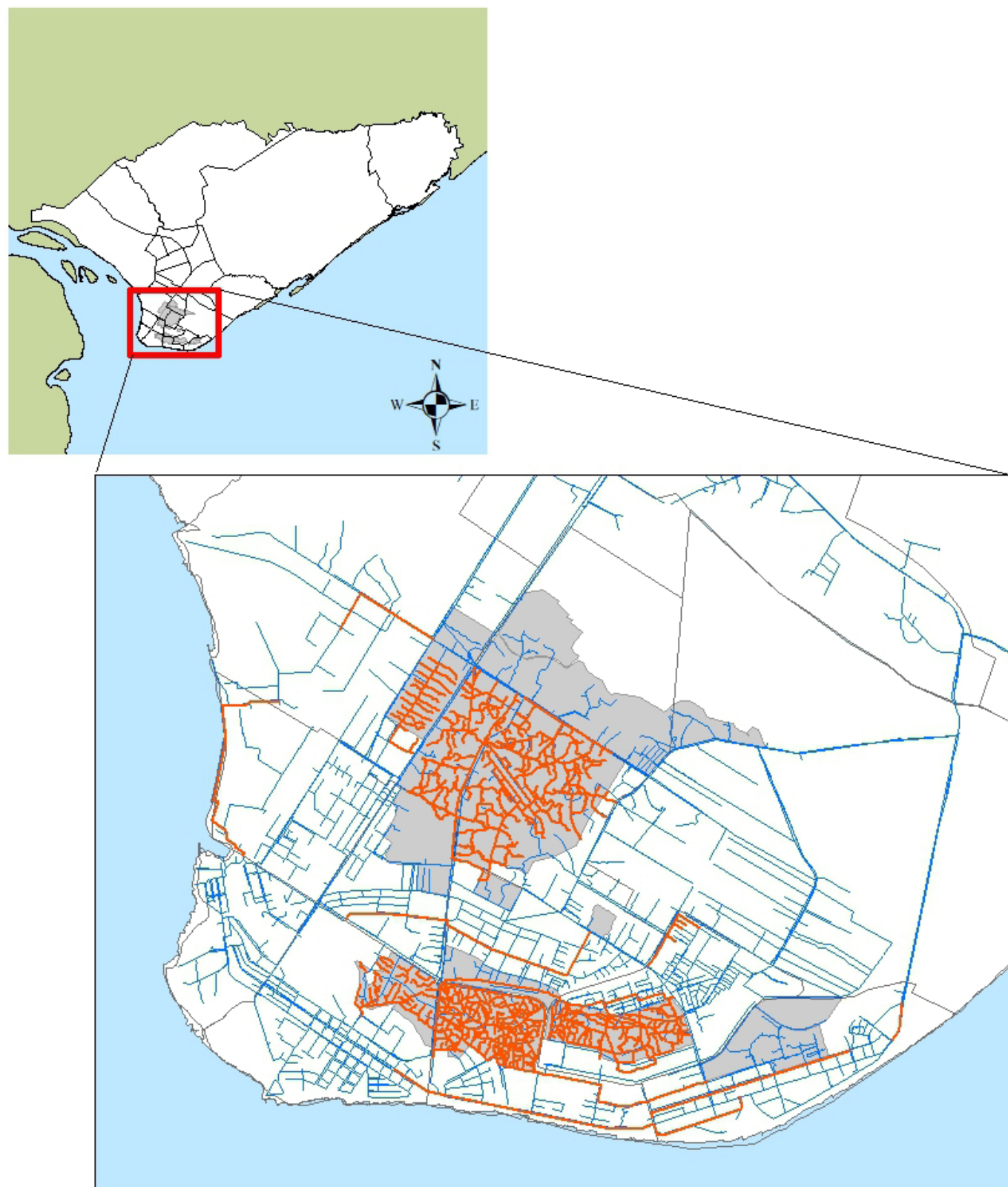
Figure 2 Map of PAASIM study site in Beira, Mozambique. Map of Beira, Mozambique, with enlargement highlighting study site. Red lines indicate the new distribution system water network. Blue lines indicate other parts of the water network. Grey shaded areas indicate neighbourhoods enrolled in the study.

of enrolment (2020) and before the end of the study (2023). FIPAG provided maps and timelines for construction works related to the upgrades, and the specific areas

participating in the water loss reduction project. We also worked with FIPAG and through satellite imagery to identify similarly dense low-income areas in Beira that were
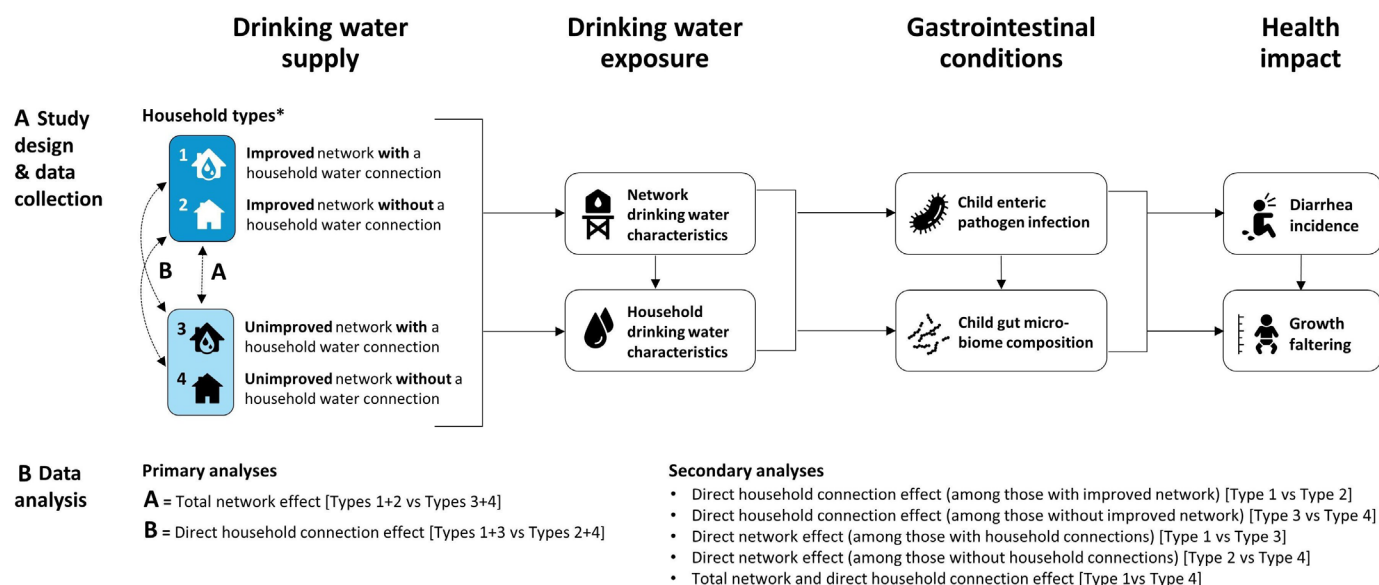
Figure 3 PAASIM study summary diagram. The diagram reflects the summarised (A) study design and data collection, and (B) data analysis approaches to isolate the effects of both overall water supply infrastructural improvements as well as the presence of a household water connection. *The dark blue shading indicates neighbourhoods that benefitted from water distribution system improvements ('improved water network') and the light blue shading indicates similar neighbourhoods that have not received improvements ('unimproved water network'). Within each neighbourhood, some households have a direct household connection to the water network and others do not.

not slated to receive water network upgrades. A total of 17 potential neighbourhoods were considered for inclusion in the study, and neighbourhoods were divided into 80 sub-neighbourhoods, delineated along natural boundaries such as roads or waterways. 'Intervention' sub-neighbourhoods include areas with the upgraded water distribution system. 'Control' sub-neighbourhoods include areas not receiving these improvements during the time period of the study. Within both intervention and control sub-neighbourhoods, some households have a connection to the water system and others do not. We excluded nine control sub-neighbourhoods that were in close proximity to intervention sub-neighbourhoods or that were scheduled to receive the interventions within the timetable of our project; some control sub-neighbourhoods are slated to receive the intervention after completion of our study.

Matching and restriction

We followed the suggestion of Arnold *et al* to use baseline (preintervention) data at the community level to match intervention to control communities when randomisation is not possible.[23] To characterise sub-neighbourhoods for further matching and restriction, we performed a population-based community survey in November–December 2020 of approximately 1700 households; this provided approximately a 5% proportional sample of our potential study sub-neighbourhoods. We used a random grid sampling approach to estimate household density, using Google Earth satellite imagery, where a grid was placed over an area, and a random selection of squares were selected and counted independently in duplicate, and the number of houses per unit was extrapolated

across unsampled squares. The survey contained modules regarding household demographics, water access and practices, sanitation access and practices, household assets and wealth indicators, as well as questions related to COVID-19. A socioeconomic status (SES) score was constructed using the 'simple poverty scorecard'[78] developed specifically for Mozambique, and scores were aggregated at the sub-neighbourhood level, and categorised into tertiles.

We matched intervention sub-neighbourhoods to control sub-neighbourhoods, using coarsened exact matching,[79 80] with intervention sub-neighbourhoods being matched to control sub-neighbourhoods within the same tertile of both SES and population density. Four neighbourhoods (encompassing nine sub-neighbourhoods) were found to be outliers in terms of their sub-neighbourhood-level SES or sanitation, and were excluded from the study sampling frame. Ultimately, we designated 36 intervention sub-neighbourhoods, with an estimated 16 800 households, and 26 control sub-neighbourhoods, with an estimated 9500 households.

Participant recruitment, eligibility and retention

We recruit pregnant women at the last trimester of pregnancy and follow the infant–mother dyads until the child is 12 months old (figure 5). We selected the first 12 months of life because it is a critical development window,[81–83] it is a time when children are most at risk of acute and chronic effects of enteropathogen infection[84] and it is a short enough period of time to avoid changes in water access that might occur. We recruit mothers at the end of their pregnancy so we can collect data on household risk factors (including drinking water quality)
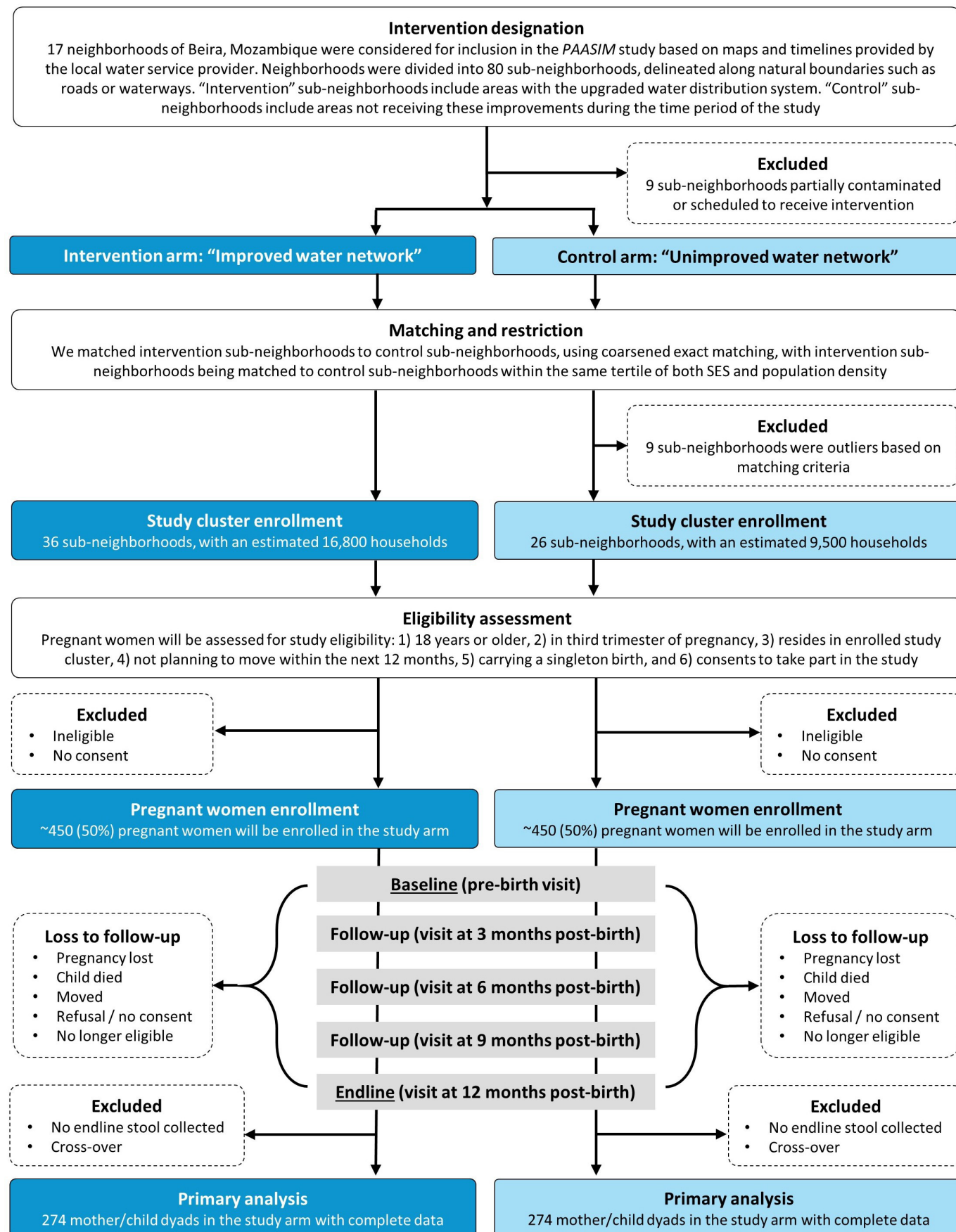
Intervention designation
17 neighborhoods of Beira, Mozambique were considered for inclusion in the *PAASIM* study based on maps and timelines provided by the local water service provider. Neighborhoods were divided into 80 sub-neighborhoods, delineated along natural boundaries such as roads or waterways. "Intervention" sub-neighborhoods include areas with the upgraded water distribution system. "Control" sub-neighborhoods include areas not receiving these improvements during the time period of the study

Excluded
9 sub-neighborhoods partially contaminated or scheduled to receive intervention

Intervention arm: "Improved water network"

Control arm: "Unimproved water network"

Matching and restriction
We matched intervention sub-neighborhoods to control sub-neighborhoods, using coarsened exact matching, with intervention sub-neighborhoods being matched to control sub-neighborhoods within the same tertile of both SES and population density

Excluded
9 sub-neighborhoods were outliers based on matching criteria

Study cluster enrollment
36 sub-neighborhoods, with an estimated 16,800 households

Study cluster enrollment
26 sub-neighborhoods, with an estimated 9,500 households

Eligibility assessment
Pregnant women will be assessed for study eligibility: 1) 18 years or older, 2) in third trimester of pregnancy, 3) resides in enrolled study cluster, 4) not planning to move within the next 12 months, 5) carrying a singleton birth, and 6) consents to take part in the study

Excluded
- Ineligible
- No consent

Excluded
- Ineligible
- No consent

Pregnant women enrollment
~450 (50%) pregnant women will be enrolled in the study arm

Pregnant women enrollment
~450 (50%) pregnant women will be enrolled in the study arm

Baseline (pre-birth visit)

Follow-up (visit at 3 months post-birth)

Follow-up (visit at 6 months post-birth)

Follow-up (visit at 9 months post-birth)

Endline (visit at 12 months post-birth)

Loss to follow-up
- Pregnancy lost
- Child died
- Moved
- Refusal / no consent
- No longer eligible

Loss to follow-up
- Pregnancy lost
- Child died
- Moved
- Refusal / no consent
- No longer eligible

Excluded
- No endline stool collected
- Cross-over

Excluded
- No endline stool collected
- Cross-over

Primary analysis
274 mother/child dyads in the study arm with complete data

Primary analysis
274 mother/child dyads in the study arm with complete data

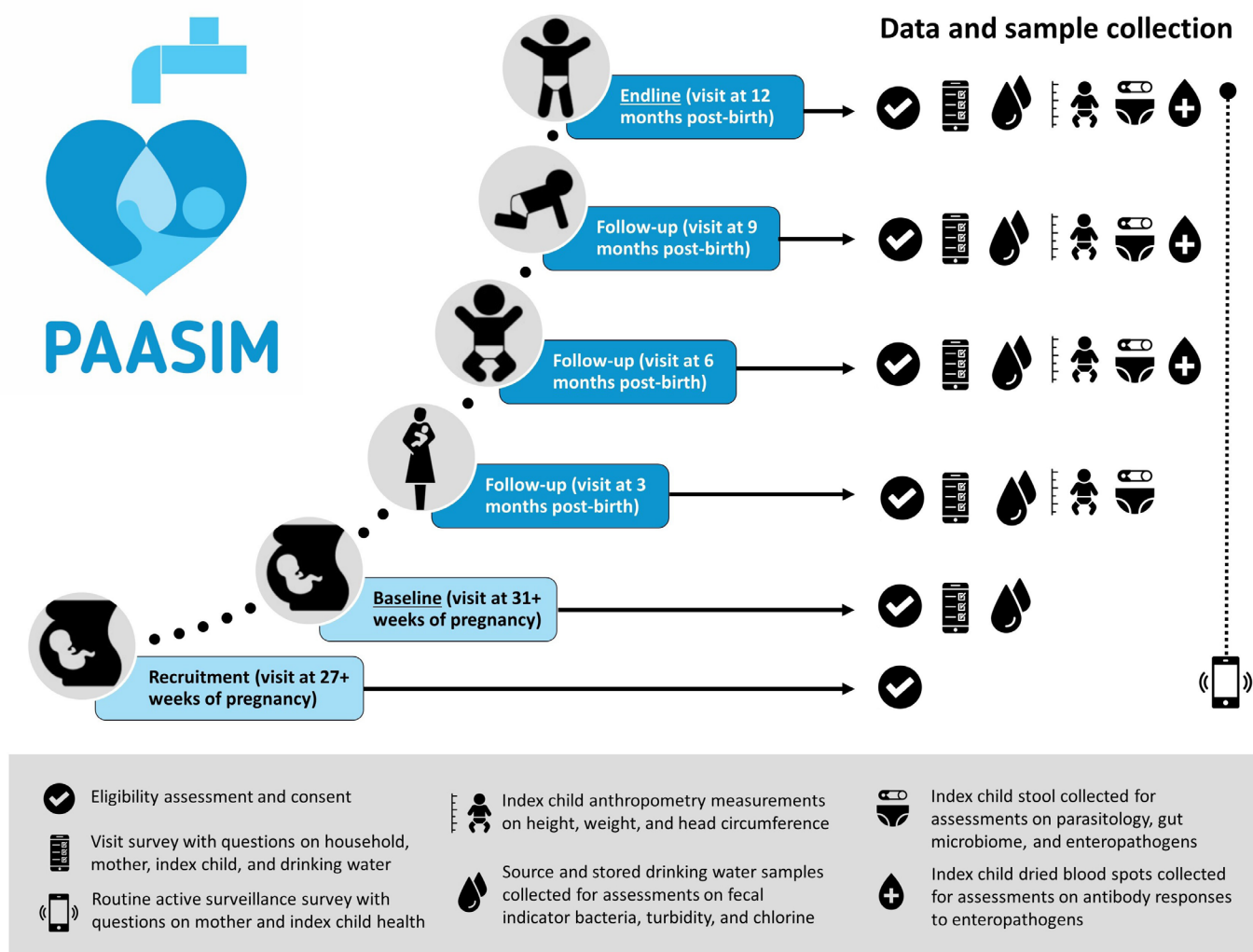**Figure 4** PAASIM study flow diagram. SES, socioeconomic status.

Figure 5 Data and sample collection timeline for outcomes in the PAASIM study. Data and sample collection of infant–mother dyads enrolled into the study will be used to address a series of questions about the impact of community-level water system improvements on acute and chronic health outcomes in children in a low-income urban area of Mozambique.

during the gestational period. Active recruitment occurs through identification of pregnant women in the 2020 population-based survey, lists of pregnant women visiting local health centres for prenatal care and study staff visiting under-enrolled sub-neighbourhoods throughout the recruitment period. Based on Ministry of Health data for Sofala Province (where Beira is located), virtually all mothers attend prenatal clinical visits.[85] Passive strategies include referrals of pregnant women by study participants and community leaders. We aim to have complete data on a total of 548 infant–mother dyads, approximately evenly divided between the intervention and control groups. We will continue to enrol dyads into both arms until we reach a minimum of 274 dyads with complete data in each arm, to ensure temporal balance throughout the duration of the study period.

During an initial pre-birth visit, pregnant women are assessed for study eligibility: (1) 18 years or older, (2) in third trimester of pregnancy, (3) resides in enrolled study cluster, (4) not planning to move within the next 12 months, (5) carrying a singleton birth and (6) consents

to take part in the study. We will reassess study eligibility at each follow-up visit and record if enrolled participants have been lost to follow-up.

Data collection
A local data collection firm (WE Consult) performs the in-country coordination of participant enrolment, data collection and sample collection. Enumerators conduct household visits before birth for consent, eligibility and conditions. At months 3, 6, 9 and 12 we deploy survey instruments to collect data on key indicators through structured observations, reports from respondents and objective measurements (online supplemental table S1). We assess a number of variables related to drinking water, including aspects of water quality, water access, water availability, water security, water consumption and participant satisfaction with water. Enumerators also conduct brief active surveillance calls on a monthly-basis by phone with caregivers to gather supplemental information on prenatal and perinatal environmental exposures and illnesses, on child illness symptoms and intake

of medicines, vitamins, breast feeding and introduction of complementary foods (figure 5). To facilitate communication with the study team, participants receive a 150 MZN (Mozambican metical) phone credit at each visit. Aside from these phone credits, there is no financial incentive provided to participants to partake in the study, per Mozambican guidelines for human subjects research. We ask the caregiver to report diarrhoea and blood in the stool (dysentery) of the index child in the previous week at the 3, 6, 9 and 12-month surveys and during active surveillance calls; due to concerns about reporting biases, we also include negative control outcomes.[86] At each post-birth visit we measure child: (1) length, weight and head circumference, and (2) calculate length-for-age and weight-for-age Z-scores. Prevalence of stunting and underweight are defined as two SD below median of the reference population.[87] All data are collected on electronic tablets using Open Data Kit Collect, an open-source programme which allows offline data collection on a mobile device.[88] Additional details are provided in the online supplemental material.

Sample collection, processing and analysis

We briefly describe sample collection and downstream processing and analysis here, with additional details provided in the online supplemental material.

Stool

Stool of the index child is collected at months 3, 6, 9 and 12. Three aliquots are placed in temperature stable lysis buffer collection tubes, and two additional aliquots are used to prepare a slide for Kato-Katz analysis of parasite ova.[89] Eligible participants are referred for deworming medicine at the 12-month visit, after returning results of the parasitological examination to study subjects in collaboration with Instituto Nacional de Saúde (INS) staff in Beira.

Extracted nucleic acids are analysed: (1) using the TaqMan Array Card (TAC, Thermo Fisher Scientific, Waltham, Massachusetts, USA) assay, which allows quantification by real-time PCR via a 384-well microfluidic card for simultaneous detection of multiple viral, bacterial and parasitic enteric pathogen targets as well as antimicrobial resistance genes,[90] customised for our targets of interest (online supplemental table S3); and (2) by sequencing of the V4 region of the 16S ribosomal RNA (rRNA) gene amplicon to characterise gut microbiome community structure and composition. Bioinformatic analyses will be completed using the QIIME2 software platform.[91]

Dried blood spots
Sample collection

A trained nurse from INS collects up to six dried spots of capillary blood of the index child at 6, 9 and 12-month visits on Tropbio Filter Paper Blood Collection Disks (Cellabs, Sydney, Australia), using a 2 mm lancet. Samples are stored at −20°C at INS facilities in Beira and shipped at ambient temperature.[92] We will use the Luminex platform to carry out high throughput, multiplex antibody assays that enable the simultaneous measurement of quantitative antibody responses to dozens of pathogens from a single blood spot.[93] Our first measure will occur at 6 months, to avoid detection of maternal antibodies that wane over the first 3–6 months of life.[94]

Drinking water
Sample collection

We collect 100 mL household drinking water samples from source and stored water at all household visits. To complement the household sampling, we collect samples from a selection of 45 public standpipes located within the study area and 55 additional public standpipes located elsewhere in the city of Beira. At public standpipes we also measure water pressure by measuring time to fill a fixed volume (1 L or 5 L, depending on the pressure). Samples are processed for faecal indicator bacteria within 6 hours of collection using Colilert-18 reagent and the Quanti-Tray/2000 MPN method (IDEXX Laboratories, Westbrook, Maine, USA), as well as for free and total chlorine levels and additional physiochemical parameters (pH, conductivity and turbidity). Large volume samples will be collected from a subset of 50 households (1-2 L, processed by membrane filtration) and 25 public standpipes (50 L, processed by dead end ultrafiltration[95]) in two different seasons, and tested for enteropathogens using the TAC assay.

Outcomes

Our primary outcomes include: any bacteria or protozoa infection at age 12 months after birth; individual pathogens or pathogen groups; child gut microbiome composition; and household source water quality. We include viral, protozoal and bacterial pathogens responsible for the vast majority of enteric pathogen infections and global disease burden.[96 97] While we measure viral pathogens using the TAC assay, they will be excluded from the combined enteropathogen prevalence primary outcome measure, because waterborne transmission is unlikely to dominate for these viral pathogens.[98–101] In addition to the aforementioned reasons related to child development and infection risk, measuring pathogens at 12 months will give us the greatest power to detect a difference, given higher levels of infection at that age than in younger children. We will measure gut microbiome using 16S rRNA gene amplicon sequencing in the full sample at 12 months and in a random subset of 200 children with complete data at 3, 6 and 9 months, evenly distributed between intervention and control groups; dyads eligible for subset sampling will include those with complete stool sample collection and unchanged intervention exposure conditions. The 12-month samples will allow us to compare all study children at a common time, when all children are consuming drinking water and once the gut microbiome has become relatively established[102]; the longitudinal samples will allow for comparison of development of the microbiome over time between the two

groups. Microbiome outcomes include alpha and beta diversity metrics, and identification of enriched taxonomic groups. We also include household source water quality as a primary exposure outcome, as understanding whether exposure to microbial contaminants is altered is considered a critical aspect of evaluation of WASH projects.[71 103]

Additional non-primary outcomes include pathogen count, pathogen community similarity (measured using Jaccard Similarity Index), diarrhoea, child growth and prior enteropathogen infection (measured using serology on dried blood spot samples). We will measure additional water quality exposure measures, as well as measures of exposure to the improved water system, such as fidelity of the intervention (eg, improvements to water quantity and coverage of household taps) and receipt of the intervention by community members (eg, reductions in water insecurity, increased water use). These fidelity and uptake measures will be collected at all time points through direct observation and respondent report. Available minimal detectable effect sizes are summarised in table 1 and calculations are further detailed in the online supplemental material.

Analysis plan
The pre-analysis plan for this study is published on the Open Science Framework platform (https://osf.io/4rkn6/).

Total network effect
To assess the impact of the intervention on our primary enteric pathogen infection outcomes and water quality exposure outcome (table 1), we will use an intention-to-treat analysis approach to compare children living in intervention versus control sub-neighbourhoods, without regard to uptake/use of the intervention (ie, direct household connection on the premises). This answers the relevant policy question 'what if an improvement is delivered to an area?' We will use multivariable log-linear binomial regression models, as pathogen infection is a binary variable, and will use generalised estimating equations to account for clustering at the sub-neighbourhood level. We group matched on sub-neighbourhood-level SES and population density, using weighting to account for unequal numbers between the intervention and control areas within each matching stratum.[104] We will additionally control for household-level and individual-level confounders, including household SES, household sanitation, mother's education-level and child sex. We may adjust for additional variables if there are found to be imbalances in potential confounders in our baseline assessment. We hypothesise that the intervention will lead to reductions in enteric pathogens among children and microbial water contamination of source water.

For additional outcomes and exposure variables of interest (table 1), we will use a similar modelling approach, using log-linear binomial regression models for binary outcomes, linear regression models for continuous outcomes and Poisson (or negative binomial) models for count outcomes. For outcomes measured at multiple time points, we will present results separately for each given time point. For these analyses, we will control for sub-neighbourhood-level SES and population density through matching, and will additionally control for household sanitation, mother's education-level, child sex and any other variables that are imbalanced and are conceivably potential confounders. For previous entero-pathogen exposure evaluated using serological measures we hypothesise that those in the intervention group will show delays in pathogen acquisition.

To assess the impact of the intervention on microbiome outcomes, we will evaluate alpha diversity (Chao1 species richness estimator, Pielou's J evenness estimator and the Shannon Diversity Index[105]) using the same modelling approach as described above for continuous outcomes. Linear discriminate effect size analyses will be used to evaluate specific 16S rRNA gene-based Operational Taxonomic Units (OTUs) that differ between individuals in intervention versus control groups, and will include effect size corrections.[106] We will examine the impact of intervention groups, controlling for other covariates, on community similarity using Adonis permutation models,[107] based on weighted UniFrac and Bray-Curtis distances, and evaluate and visualise differences using principal components analysis (PCA) and/or non-metric multidimensional scaling (NMDS) plots. We hypothesise that we will be able to observe detectable differences in gut microbiome composition in children living in intervention versus control sub-neighbourhoods and we will report these differences at the individual OTU and bacterial family levels.

Direct household connection effect
To assess the effect of having a water connection at the household or compound, we will use models similar to those described above, but accounting for a household network connection. We will also assess the interaction between the household and neighbourhood network variables, which will allow us to contrast and estimate indirect, direct and total effects, as shown in figure 3. We hypothesise that participants with both improved water networks in their sub-neighbourhoods and household water connections will most benefit from the interventions in terms of our primary and non-primary health outcomes and exposure outcomes of interest.

Additional analyses
For select primary outcomes, we will assess if there is effect modification by a third variable, such as follow-up round/age, participant sex and household sanitation access. We will use interaction terms to identify potential interactions, and will present stratified results (eg, separately by sex) if interactions are detected. The intervention status of sub-neighbourhoods was set at baseline, but if control sub-neighbourhood(s) receive the intervention after the study has started, we will perform sensitivity analyses

Table 1 Primary and non-primary health outcomes and exposure outcomes for the PAASIM study

Primary outcomes	Time point	Minimum detectable effect (risk ratio)	Anticipated control group prevalence*
Patent enteric pathogen infection			
Prevalence of any bacterial or protozoan pathogen	12 months†	0.74	70%
Prevalence of any bacterial pathogen	12 months†	0.69	61%
Prevalence of any protozoan pathogen	12 months†	0.50	32%
Any co-infection (bacterial, protozoan or viral pathogens)	12 months†	0.60	45%
Gut microbiome composition	12 months‡		
Alpha diversity		–	NA
Beta diversity		–	NA
Enriched taxa		–	NA
Household source drinking water quality	12 months†	–	NA
Additional health outcomes			
Patent enteric pathogen infection			
Pathogen count	12 months†	–	–
Pathogen community similarity	12 months†	–	–
Individual pathogens	12 months†	n/a–0.49	2–31%
Any virus	12 months†	0.46	28%
Gut microbiome composition	3, 6, 9 months‡		
(Same variables as gut microbiome composition for 12 months)			
Prior enteric pathogen exposure	6, 9, 12 months	–	–
(Same variables as patent enteric pathogen infection)			
Diarrhoea 1-week period prevalence (caregiver report)	Weekly§	0.26	14.4%
Anthropometric measurements	3, 6, 9, 12 months		
Length-for-age Z-score		–	–
Weight-for-age Z-score		–	–
Stunting prevalence¶		0.49	31%
Underweight prevalence¶		0.22	13%
All-cause mortality (while enrolled in study)	Continuous	–	–
Additional exposure outcomes			
Primary drinking water source	3, 6, 9, 12 months		
Drinking water quality (source)	3, 6, 9 months**		
Drinking water quality (stored)	3, 6, 9, 12 months		
Water access	3, 6, 9, 12 months		
Water availability	3, 6, 9, 12 months		
Water security	3, 6, 9, 12 months		
Water consumption	3, 6, 9, 12 months		
User satisfaction with water	3, 6, 9, 12 months		

Table 1 Continued

Primary outcomes	Time point	Minimum detectable effect (risk ratio)	Anticipated control group prevalence*
Patent enteric pathogen infection in stool is measured via TaqMan Array Card assay; Stool microbiome composition is measured via 16s ribosomal RNA amplicon sequencing; prior enteric pathogen exposure is measured via serological assays of dried blood spots. Drinking water quality measured by IDEXX as *Escherichia coli* most probable number/100 mL. See text for further details. Calculations and additional values for the minimum detectable effect are described in the online supplemental material.			

*Anticipated control group prevalence based on control group prevalence for 10–14 months old in the MapSan trial.[32]
† Samples also collected at 3, 6 and 9 months of age may also be analysed, depending on results of primary analysis at 12 months.
‡A subset of n=200 samples will be analysed for gut microbiome composition in children at 3, 6 and 9 months of age. All 12-month samples will be analysed.
§This is a conservative estimate as it does not account for weekly active surveillance.
¶Defined as two SD below median of the reference population.
**Samples will be analysed at 3, 6, 9 and 12 months, but 12-month samples are the primary outcome of interest.

dropping and/or recategorising sub-neighbourhood(s) that crossed over.

There are several analyses where we do not use the matched-design and intervention variable in our analyses. For example, we will assess associations between various water measures on health, without regard to the intervention designation. We will also examine changes in the gut microbiomes of children over time. Additional analyses will be described and documented in OSF.

Sample size and power calculations

Our minimal sample size of 548 households—half in intervention and half in matched control sub-neighbourhoods—was powered for our primary outcome of prevalence of any non-viral pathogen. Using data from the MapSan trial for children 10–14 months of age (J. Knee, *pers comm*) we used a control group prevalence of 70% for any non-viral pathogen, and estimated the ability to detect a relative risk of 0.74, alpha=0.05 and power=80% using a two-sided test for significance.[32] We estimated a sub-neighbourhood-level interclass correlation coefficient (ICC) of 0.05 (a moderate estimate) among our 62 designated sub-neighbourhoods. We will also report on the final ICC and other assumptions of this power analysis at the end of the study. Estimates of minimum detectable effect sizes based on control prevalence of the outcome of interest (online supplemental table S2) show we may be adequately powered to detect a difference in some individual pathogens if those pathogens have high prevalence and/or if they are strongly associated with the water supply improvement intervention (eg, waterborne pathogens). We target planned recruitment at 900 pregnant women in the third trimester, to account for incomplete data and loss to follow-up. We used sub-neighbourhood enrolment targets proportionate to our density estimates to achieve balance across intervention and control sub-neighbourhoods.

Blinding

All laboratory personnel and field enumerators are blinded to the intervention status of the samples and households. Participants cannot be blinded to their household-level water exposure status or cluster-level exposure status, although participants may or may not know about water improvements in their particular neighbourhood. A primary analyst external to the core data management team is blinded to the group assignments until the data cleaning and primary analysis are completed. Details of these procedures are included in the online supplemental material. Unblinding will occur only after primary outcome models are developed and compared between two independent analysts. Analyses examining the impact of the intervention on non-primary outcomes or exposures of interest will not be unblinded until after analyses that examine the impact of the intervention on our primary outcomes have been completed. Purely observational analyses that do not require information on intervention groups may be completed before unblinding occurs.

ETHICS AND DISSEMINATION

The study protocol, informed consent forms and data collection tools were approved by (1) Mozambique National Bio-Ethics Committee for Health (IRB00002657) and (2) Emory University's Institutional Review Board (IRB00098584). Prior to enrolment, study staff fully explained and carried out the consent process and documented the procedure. Subjects provided written consent with a signature. In the case of illiteracy of the subject, study staff verbally summarised the material with the subject, and the participants were required to provide written consent by marking the document with a thumbprint. As this study is a natural experiment that the investigators do not control, we do not have a data monitoring committee or any interim stopping guidelines. Enrolment for this study began during the COVID-19 pandemic, and precautions were taken to secure the safety of study staff and participants based on guidance from INS, Emory University, and the University of Washington.

Any changes to this published protocol will be noted in OSF, and, where relevant, in future publications. De-identified data sufficient to replicate study findings will be

publicly available on OSF on completion and publication of the study results. A report will also be prepared and shared with the municipality and health authorities in Beira, and other relevant stakeholders. All microbial DNA sequence data will be made available through the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) on validation and/or publication of the corresponding manuscript.

Author affiliations
[1]Department of Environmental and Occupational Health Sciences, University of Washington School of Public Health, Seattle, Washington, USA
[2]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[3]Division of Biostatistics, Epidemiology and Environmental Health, School of Public Health, University of Nevada Reno, Reno, Nevada, USA
[4]WE Consult, Maputo, Mozambique
[5]Beira Operations Research Center, National Health Institute (INS), Ministry of Health of Mozambique, Beira, Mozambique
[6]Autoridade Reguladora de Água, Instituto Público, Maputo, Mozambique
[7]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[8]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA
[9]Environmental Science and Engineering, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, North Carolina, USA
[10]Ministry of Health, Instituto Nacional de Saúde, Marracuene, Mozambique

Twitter Karen Levy @klevy_uw, Bacelar Muneme @BacelarMuneme and Matthew C Freeman @MatthewCFreeman

Contributors MM and MCF conceived of the overarching idea of evaluating the intervention. KL and MCF conceived of the specific study and secured funding. JVG and MCF designed the analysis plan, with input from LAW. ZAC, AJ, BM, SH, SM, JSS and MKM-P designed protocols for recruitment of participants and oversaw collection of field data, with guidance from RN and JLM. JSS and JVG oversaw data management, with help from SH, MKM-P, SM and CSF-S. CV and CSF-S designed specimen management and laboratory protocols. TFC, JB and LAW advised on study design, epidemiological approaches and research methods. RN, MM and JLM provided oversight on relevant scientific questions in Mozambique. RN, JSS, CV, KL, MCF, ZAC, SM and MKM-P managed human subjects protocol submissions. KTK and KL oversaw microbiome analysis approach. RN oversaw parasitology analysis approach. KL and JB oversaw enteric pathogen analysis approach. KL and JVG wrote the manuscript, with input from all authors.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographical or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Karen Levy http://orcid.org/0000-0002-0968-9401
Christine S Fagnant-Sperati http://orcid.org/0000-0001-9794-307X
Molly K Miller-Petrie http://orcid.org/0000-0001-7493-5915
Lance A Waller http://orcid.org/0000-0001-5002-8886
Joe Brown http://orcid.org/0000-0002-5200-4148
Matthew C Freeman http://orcid.org/0000-0002-1517-2572

REFERENCES
1 Centers for Disease Control and Prevention. Achievements in public health, 1900-1999: control of infectious diseases. *MMWR Morb Mortal Wkly Rep* 1999;48:621–9.
2 Cutler D, Miller G. The role of public health improvements in health advances: the Twentieth-century United States. *Demography* 2005;42:1–22.
3 Prüss-Ustün A, Bartram J, Clasen T, et al. Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. *Trop Med Int Health* 2014;19:894–905.
4 Global Burden of Disease Collaborative Network. Global burden of disease study(GBD 2019) under-5 mortality and adult mortality 1950-2019. Seattle, USA: Institute for Health Metrics and Evaluation (IHME), 2020.
5 Victora CG, Christian P, Vidaletti LP, et al. Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda. *Lancet* 2021;397:1388–99.
6 Danaei G, Andrews KG, Sudfeld CR, et al. Risk factors for childhood stunting in 137 developing countries: a comparative risk assessment analysis at global, regional, and country levels. *PLoS Med* 2016;13:e1002164.
7 World Bank/Interational Monitary Fund (IMF). *Global montioring report 2013: rural-urban dynamics and the millennium development goals.* Washington, DC: World Bank/IMF, 2013.
8 United Nations Department of Economic and Social Affairs. *World urbanization prospects: the 2018 revision.* United Nations, 2019.
9 United Nations-HABITAT. The state of african cities: re-imagining sustainable urban transitions. 2014. Available: http://unhabitat.org/books/state-of-african- cities-2014-re-imagining-sustainable-urban-transitions/
10 Jones O, Mansour G, Burr P. *The of wash financing in eastern and Southern Africa regional level assessment.* Nairobi: UNICEF Eastern and Southern Africa Regional Office, 2019.
11 Bain R, Cronk R, Wright J, et al. Fecal contamination of drinking-water in low- and middle-income countries: a systematic review and meta-analysis. *PLOS Med* 2014;11:e1001644.
12 Hunter PR, Zmirou-Navier D, Hartemann P. Estimating the impact on health of poor reliability of drinking water interventions in developing countries. *Sci Total Environ* 2009;407:2621–4.
13 World Health Organization. GLAAS 2018/2019 country survey. 2018. Available: https://www.who.int/publications/m/item/glaas-2018-2019-country-survey
14 United Nations. Investing in water and sanitation: increasing access, reducing inequalities GLAAS 2014 findings-- special report for africa. UN water & world health organization 2015. n.d. Available: http://www.who.int/water_sanitation_health/glaas/glaas2014-africa-region.pdf
15 UN-Water. UN-water GLAAS 2019: national systems to support drinking-water, sanitation and hygiene - global status report 2019. 2019. Available: https://www.unwater.org/publications/un-water-glaas-2019-national-systems-to-support-drinking-water-sanitation-and-hygiene-global-status-report-2019/
16 Ercumen A, Arnold BF, Kumpel E, et al. Upgrading a piped water supply from intermittent to continuous delivery and association with waterborne illness: a matched cohort study in urban India. *PLOS Med* 2015;12:e1001892.

17 Pickering AJ, Crider Y, Sultana S, *et al*. Effect of in-line drinking water chlorination at the point of collection on child diarrhoea in urban bangladesh: a double-blind, cluster-randomised controlled trial. *Lancet Glob Health* 2019;7:e1247–56.

18 Reese H, Routray P, Torondel B, *et al*. 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. *Int J Epidemiol* 2019;48:1757–67.

19 Clasen TF, Alexander KT, Sinclair D, *et al*. Interventions to improve water quality for preventing diarrhoea. *Cochrane Database Syst Rev* 2015;2015:CD004794.

20 Wolf J, Hubbard S, Brauer M, *et al*. Effectiveness of interventions to improve drinking water, sanitation, and handwashing with soap on risk of diarrhoeal disease in children in low-income and middle-income settings: a systematic review and meta-analysis. *Lancet* 2022;400:48–59.

21 Sanson-Fisher RW, Bonevski B, Green LW, *et al*. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med* 2007;33:155–61.

22 Spears D, Ban R, Cumming O. Trials and tribulations: the rise and fall of the RCT in the WASH sector. In: *Randomized control trials in the field of development*. Oxford: Oxford University Press, 2020.

23 Arnold BF, Khush RS, Ramaswamy P, *et al*. Causal inference methods to study nonrandomized, preexisting development interventions. *Proc Natl Acad Sci U S A* 2010;107:22605–10.

24 Boisson S, Schmidt W-P, Berhanu T, *et al*. Randomized controlled trial in rural Ethiopia to assess a portable water treatment device. *Environ Sci Technol* 2009;43:5934–9.

25 Luoto J, Mahmud M, Albert J, *et al*. Learning to dislike safe water products: results from a randomized controlled trial of the effects of direct and peer experience on willingness to pay. *Environ Sci Technol* 2012;46:6244–51.

26 Levy K, Nelson KL, Hubbard A, *et al*. Following the water: a controlled study of drinking water storage in northern coastal Ecuador. *Environ Health Perspect* 2008;116:1533–40.

27 Wright J, Gundry S, Conroy R. Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. *Trop Med Int Health* 2004;9:106–17.

28 Levy K, Anderson L, Robb KA, *et al*. Household effectiveness vs. laboratory efficacy of point-of-use chlorination. *Water Res* 2014;54:69–77.

29 Stauber CE, Kominek B, Liang KR, *et al*. Evaluation of the impact of the plastic biosand filter on health and drinking water quality in rural tamale, Ghana. *Int J Environ Res Public Health* 2012;9:3806–23.

30 Brown J, Clasen T. High adherence is necessary to realize health gains from water quality interventions. *PLoS One* 2012;7:e36735.

31 Enger KS, Nelson KL, Rose JB, *et al*. The joint effects of efficacy and compliance: a study of household water treatment effectiveness against childhood diarrhea. *Water Res* 2013;47:1181–90.

32 Knee J, Sumner T, Adriano Z, *et al*. Effects of an urban sanitation intervention on childhood enteric infection and diarrhea in maputo, Mozambique: a controlled before-and-after trial. *Elife* 2021;10:e62278.

33 Gough EK, Moulton LH, Mutasa K, *et al*. Effects of improved water, sanitation, and hygiene and improved complementary feeding on environmental enteric dysfunction in children in rural Zimbabwe: a cluster-randomized controlled trial. *PLoS Negl Trop Dis* 2020;14:e0007963.

34 Null C, Stewart CP, Pickering AJ, *et al*. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural kenya: a cluster-randomised controlled trial. *Lancet Glob Health* 2018;6:e316–29.

35 Luby SP, Rahman M, Arnold BF, *et al*. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural bangladesh: a cluster randomised controlled trial. *Lancet Glob Health* 2018;6:e302–15.

36 Post A, Ray I. Hybrid modes of urban water delivery in low- and middle-income countries. *Oxf Res Encycl Environ Sci* 2020.

37 World Health Organization (WHO) and the United Nations Children's Fund (UNICEF). *Progress on household drinking water, sanitation and hygiene 2000-2020: five years into the sdgs*. Geneva: World Health Organization (WHO) and the United Nations Children's Fund (UNICEF, 2021.

38 Walker CLF, Rudan I, Liu L, *et al*. Global burden of childhood pneumonia and diarrhoea. *Lancet* 2013;381:1405–16.

39 Liu L, Johnson HL, Cousens S, *et al*. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* 2012;379:2151–61.

40 Arnold BF, Galiani S, Ram PK, *et al*. Optimal recall period for caregiver-reported illness in risk factor and intervention studies: a multicountry study. *Am J Epidemiol* 2013;177:361–70.

41 Baqui AH, Black RE, Yunus M, *et al*. Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. *Int J Epidemiol* 1991;20:1057–63.

42 Wright J, Gundry SW, Conroy RM. A review of changing episode definitions and their effects on estimates of diarrhoeal morbidity. *J Health Popul Nutr* 2007;25:448–55.

43 Wright JA, Gundry SW, Conroy R, *et al*. Defining episodes of diarrhoea: results from a three-country study in sub-Saharan Africa. *J Health Popul Nutr* 2006;24:8–16.

44 Zafar SN, Luby SP, Mendoza C. Recall errors in a Weekly survey of diarrhoea in Guatemala: determining the optimal length of recall. *Epidemiol Infect* 2010;138:264–9.

45 Trostle JA, Yépez-Montufar JA, Corozo-Angulo B, *et al*. Diarrheal illnesses on the ecuadorian coast: socio-environmental changes and health concepts. *Cad Saude Publica* 2010;26:1334–44.

46 Blum D, Feachem RG. Measuring the impact of water supply and sanitation investments on diarrhoeal diseases: problems of methodology. *Int J Epidemiol* 1983;12:357–65.

47 Lee G, Cama V, Gilman RH, *et al*. Comparison of two types of epidemiological surveys aimed at collecting daily clinical symptoms in community-based longitudinal studies. *Ann Epidemiol* 2010;20:151–8.

48 Schmidt W-P, Arnold BF, Boisson S, *et al*. Epidemiological methods in diarrhoea studies -- an update. *Int J Epidemiol* 2011;40:1678–92.

49 Thapar N, Sanderson IR. Diarrhoea in children: an interface between developing and developed countries. *Lancet* 2004;363:641–53.

50 Holcomb DA, Knee J, Sumner T, *et al*. Human fecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in maputo, mozambique. *Int J Hyg Environ Health* 2020;226:113496.

51 Chard AN, Baker KK, Tsai K, *et al*. Associations between soil-transmitted helminthiasis and viral, bacterial, and protozoal enteroinfections: a cross-sectional study in rural laos. *Parasit Vectors* 2019;12:216.

52 Brown J, Cumming O, Bartram J, *et al*. A controlled, before-and-after trial of an urban sanitation intervention to reduce enteric infections in children: research protocol for the maputo sanitation (mapsan) study, mozambique. *BMJ Open* 2015;5:e008215.

53 Liu J, Kabir F, Manneh J, *et al*. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study. *Lancet Infect Dis* 2014;14:716–24.

54 Lunn PG, Northrop-Clewes CA, Downes RM. Intestinal permeability, mucosal injury, and growth faltering in Gambian infants. *Lancet* 1991;338:907–10.

55 Humphrey JH. Child undernutrition, tropical enteropathy, toilets, and handwashing. *Lancet* 2009;374:1032–5.

56 Guerrant RL, Oriá RB, Moore SR, *et al*. Malnutrition as an enteric infectious disease with long-term effects on child development. *Nutr Rev* 2008;66:487–505.

57 Campbell DI, Elia M, Lunn PG. Growth faltering in rural Gambian infants is associated with impaired small intestinal barrier function, leading to endotoxemia and systemic inflammation. *J Nutr* 2003;133:1332–8.

58 Goto R, Mascie-Taylor CGN, Lunn PG. Impact of intestinal permeability, inflammation status and parasitic infections on infant growth faltering in rural Bangladesh. *Br J Nutr* 2009;101:1509–16.

59 Lunn PG. Growth retardation and stunting of children in developing countries. *Br J Nutr* 2002;88:109–10.

60 Guerrant RL, DeBoer MD, Moore SR, *et al*. The impoverished gut -- a triple burden of diarrhoea, stunting and chronic disease. *Nat Rev Gastroenterol Hepatol* 2013;10:220–9.

61 Eppig C, Fincher CL, Thornhill R. Parasite prevalence and the worldwide distribution of cognitive ability. *Proc Biol Sci* 2010;277:3801–8.

62 Wierzba TF, Muhib F. Exploring the broader consequences of diarrhoeal diseases on child health. *Lancet Glob Health* 2018;6:e230–1.

63 Yatsunenko T, Rey FE, Manary MJ, *et al*. Human gut microbiome viewed across age and geography. *Nature* 2012;486:222–7.

64 Schnorr SL. The diverse microbiome of the hunter-gatherer. *Nature* 2015;518:S14–5.

65 Schnorr SL, Candela M, Rampelli S, *et al*. Gut microbiome of the hadza hunter-gatherers. *Nat Commun* 2014;5:3654.

66 Clemente JC, Pehrsson EC, Blaser MJ, *et al*. The microbiome of uncontacted amerindians. *Sci Adv* 2015;1:e1500183.

67 Martínez I, Stegen JC, Maldonado-Gómez MX, *et al*. The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep* 2015;11:527–38.

68 Obregon-Tito AJ, Tito RY, Metcalf J, *et al*. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun* 2015;6:6505.

69 Gacesa R, Kurilshikov A, Vich Vila A, *et al*. Environmental factors shaping the gut microbiome in a dutch population. *Nature* 2022;604:732–9.

70 Clasen T, Boisson S. Assessing the health impact of water quality interventions in low-income settings: concerns associated with blinded trials and the need for objective outcomes. *Environ Health Perspect* 2016;124:886–9.

71 Goddard FGB, Ban R, Barr DB, *et al*. Measuring environmental exposure to enteric pathogens in low-income settings: review and recommendations of an interdisciplinary Working group. *Environ Sci Technol* 2020;54:11673–91.

72 Prendergast A, Kelly P. Enteropathies in the developing world: neglected effects on global health. *Am J Trop Med Hyg* 2012;86:756–63.

73 Brown J, Cumming O. Stool-Based pathogen detection offers advantages as an outcome measure for water, sanitation, and hygiene trials. *Am J Trop Med Hyg* 2020;102:260–1.

74 INE. *Resultados definitivos censo 2017 IV recenseamento geral da populacao e habitacao de mocambique*. Maputo: INE, 2017. Available: Www.Ine.Gov.Mzww.Ine.Gov.Mz

75 Water & Sanitation for the Urban Poor. *Supporting the provision of universal water services in beira and dondo*. London: Water & Sanitation for the Urban Poor, 2022.

76 World Bank. Development projects: water services & institutional support II. n.d. Available: https://projects.worldbank.org/en/projects-operations/project-detail/P149377

77 Dutch Development results Mozambique. Netherlands M of FA of the K of the. Available: https://www.dutchdevelopmentresults.nl/2019/countries/mozambique#water [Accessed 20 Apr 2022].

78 Schreiner M. n.d. Simple poverty scorecard® poverty-assessment tool mozambique.

79 Blackwell M, Iacus S, King G, *et al*. Cem: coarsened exact matching in stata. *The Stata Journal* 2009;9:524–46.

80 Porro G, King G, Iacus S. CEM: software for coarsened exact matching. *J Stat Softw* 2009;30.

81 Mbuya MNN, Humphrey JH. Preventing environmental enteric dysfunction through improved water, sanitation and hygiene: an opportunity for stunting reduction in developing countries. *Matern Child Nutr* 2016;12:106–20.

82 Cumming O, Cairncross S. Can water, sanitation and hygiene help eliminate stunting? current evidence and policy implications. *Matern Child Nutr* 2016;12:91–105.

83 Oriá RB, Murray-Kolb LE, Scharf RJ, *et al*. Early-Life enteric infections: relation between chronic systemic inflammation and poor cognition in children. *Nutr Rev* 2016;74:374–86.

84 Platts-Mills JA, Babji S, Bodhidatta L, *et al*. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob Health* 2015;3:e564–75.

85 Direcção Provincial de Saúde. *Balanço do plano economico e social*. Beira: Governo de Sofala, Direcção Provincial de Saúde, República de Moçambique, 2021.

86 Arnold BF, Ercumen A. Negative control outcomes. *JAMA* 2016;316:2597.

87 UNICEF. Nutrition: definitions of the indicators. Available: http://www.unicef.org/infobycountry/stats_popup2.html [Accessed 5 Feb 2015].

88 Hartung C, Lerer A, Anokwa Y, *et al*. *Open data kit: tools to build information services for developing regions. in: proceedings of the 4th ACM/IEEE international conference on information and communication technologies and development - ICTD '10*. London, United Kingdom: ACM Press, 2010: 1–12.

89 Bosch F, Palmeirim MS, Ali SM, *et al*. Diagnosis of soil-transmitted helminths using the kato-katz technique: what is the influence of stirring, storage time and storage temperature on stool sample egg counts? *PLoS Negl Trop Dis* 2021;15:e0009032.

90 Liu J, Gratz J, Amour C, *et al*. A laboratory-developed TaqMan array card for simultaneous detection of 19 enteropathogens. *J Clin Microbiol* 2013;51:472–80.

91 Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37:852–7.

92 Centers for Disease Control and Prevention Division of Parasitic Diseases and Malaria. Blood collection and dried blood spot preparation. Atlanta, GA: Centers for Disease Control and Prevention, 2019.

93 Lammie PJ, Moss DM, Brook Goodhew E, *et al*. Development of a new platform for neglected tropical disease surveillance. *Int J Parasitol* 2012;42:797–800.

94 Zinkernagel RM. Maternal antibodies, childhood infections, and autoimmune diseases. *N Engl J Med* 2001;345:1331–5.

95 Smith CM, Hill VR. Dead-End hollow-fiber ultrafiltration for recovery of diverse microbes from water. *Appl Environ Microbiol* 2009;75:5284–9.

96 Vos T, Lim SS, Abbafati C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;396:1204–22.

97 Kotloff KL, Nasrin D, Blackwelder WC, *et al*. The incidence, aetiology, and adverse clinical consequences of less severe diarrhoeal episodes among infants and children residing in low-income and middle-income countries: a 12-month case-control study as a follow-on to the global enteric multicenter study (gems). *Lancet Glob Health* 2019;7:e568–84.

98 Strina A, Rodrigues LC, Cairncross S, *et al*. Factors associated with rotavirus diarrhoea in children living in a socially diverse urban centre in Brazil. *Trans R Soc Trop Med Hyg* 2012;106:445–51.

99 Chen S-Y, Chiu C-H. Worldwide molecular epidemiology of norovirus infection. *Paediatr Int Child Health* 2012;32:128–31.

100 da Silva Poló T, Peiró JR, Mendes LCN, *et al*. Human norovirus infection in latin america. *J Clin Virol* 2016;78:111–9.

101 Nasser AM. Prevalence and fate of hepatitis A virus in water. *Critic Rev Environment Sci Technol* 1994;24:281–323.

102 Derrien M, Alvarez A-S, de Vos WM. The gut microbiota in the first decade of life. *Trends Microbiol* 2019;27:997–1010.

103 Cumming O, Arnold BF, Ban R, *et al*. The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. *BMC Med* 2019;17:173.

104 Iacus SM, King G, Porro G. Matching for causal inference without balance checking. *SSRN J* 2008.

105 Magurran AE. *Measuring biological diversity*. Malden, MA: Blackwell Pub, 2004.

106 Segata N, Izard J, Waldron L, *et al*. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;12:R60.

107 Oksanen J, Blanchet FG, Kindt R, *et al*. Package 'vegan': community ecology package. 2016:ordination methods, diversity analysis and other functions for community and vegetation ecologists. Available: https://github.com/vegandevs/vegan [Accessed 5 Feb 2017].