

A Validated National Framework for Multi-Pathogen Wastewater-Based Epidemiological Surveillance in Kenya: From Pathogen Prioritisation to Real- Time Public Health Intelligence

Leonard Kingwara

leonard.kingwara@gmail.com

Kenya National Public Health Institute (KNPHI)

Rukia Madada

Kenya National Public Health Institute (KNPHI)

Harris Onywera

Africa Centres for Disease Control and Prevention

Samuel Oyola

International Livestock Research Institute

Noah Hull

Association of Public Health Laboratories

Isabella Ayagah

Kenya National Public Health Institute (KNPHI)

Victoria Kanana

Kenya National Public Health Institute (KNPHI)

Vera Onwonga

Kenya National Public Health Institute (KNPHI)

Shalyn Akasa

Kenya National Public Health Institute (KNPHI)

Susan Githii

Kenya National Public Health Institute (KNPHI)

Juma Bonventure

Centers for Disease Control and Prevention

Emmanuel Okunga

Kenya National Public Health Institute (KNPHI)

Naomi Luchi

Centers for Disease Control and Prevention

Sylvia Omulo

Washington State University

Isaac Ngere

Washington State University

Charles Rombo

Kenya National Public Health Institute (KNPHI)

Loice Wanjiru

Strathmore University

Benjamin Murkomen

Kenya National Public Health Institute (KNPHI)

Lolem Lokolile

Ministry of Health

Bernard Mware

International Livestock Research Institute

Nancy Etyang

Ministry of Health

Joanne Hassan

Kenya Medical Research Institute

Mark Nanyingi

Kenya National Public Health Institute (KNPHI)

Osborn Otieno

Global Fund to Fight AIDS, Tuberculosis and Malaria

Fatim Cham

Global Fund to Fight AIDS, Tuberculosis and Malaria

Peter Cherutich

Ministry of Health

Samoel Khamadi

Kenya Medical Research Institute

Ouma Oluga

Ministry of Health

Kamene Kimenye

Kenya National Public Health Institute (KNPHI)

Study protocol

Keywords: wastewater-based epidemiology, environmental surveillance, Kenya, multi-pathogen surveillance, antimicrobial resistance, poliovirus, metagenomics, digital PCR, TaqMan Array Card, sub-Saharan Africa

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11257307/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

A Validated National Framework for Multi-Pathogen Wastewater-Based Epidemiological Surveillance in Kenya: From Pathogen Prioritisation to Real-Time Public Health Intelligence

Leonard Kingwara^{1†}, Rukia Madada¹, Harris Onywera^{2,3}, Samuel Oyola⁴, Noah Hull⁵, Isabella Ayagah¹, Victoria Kanana¹, Vera Onwonga¹, Shalyn Akasa¹, Susan Githii¹, Juma Bonventure⁶, Emmanuel Okunga¹, Naomi Luchi⁶, Sylvia Omulo⁷, Isaac Ngere⁷, Charles Rombo¹, Loice Wanjiru⁸, Benjamin Murkomen¹, Lolem Lokolile⁹, Bernard Mware⁴, Nancy Etyang⁹, Joanne Hassan¹¹, Mark Nanyingi¹, Osborn Otieno¹⁰, Fatim Cham¹⁰, Peter Cherutich⁹, Samoel Khamadi¹¹, Ouma Oluga⁹, **Kamene Kimenye**^{1†}

Affiliations

- 1 Kenya National Public Health Institute (KNPHI), Ministry of Health, Nairobi, Kenya
- 2 Africa Centres for Disease Control and Prevention (Africa CDC), Addis Ababa, Ethiopia
- 3 Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa
- 4 International Livestock Research Institute (ILRI), Nairobi, Kenya
- 5 Association of Public Health Laboratories (APHL), Bethesda, MD, USA
- 6 U.S. Centers for Disease Control and Prevention (CDC), Nairobi, Kenya / Atlanta, GA, USA
- 7 Paul G. Allen School for Global Health, Washington State University, Pullman, WA, USA
- 8 Strathmore University, Nairobi, Kenya
- 9 Ministry of Health, Nairobi, Kenya
- 10 The Global Fund to Fight AIDS, Tuberculosis and Malaria, Geneva, Switzerland
- 11 Centre for Virus Research, Kenya Medical Research Institute (KEMRI), Nairobi, Kenya

[†]**Correspondence Author:** leonard.kingwara@gmail.com

Funding: Ministry of Health Health Emergency Preparedness, Response and Resilience (HEPRR) program; European Union; the Gates Foundation; and the United Nation Environmental Program (UNEP)

Abstract

Sub-Saharan Africa carries a disproportionate burden of global infectious diseases. However, its surveillance architecture is still reactive, clinic-centered, and blind to community-level transmission from asymptomatic and undiagnosed carriers. Wastewater-based epidemiology (WBE) addresses this blind spot by measuring the aggregate pathogen signal shed from the contributing population. Here, we describe a validated national protocol for multipathogen WBE in Kenya, delivered as an end-to-end, quality-assured framework. In the protocol, we integrate five components: (i) a structured seven-criterion pathogen prioritisation matrix, applied to a 64-pathogen candidate list drawn from Kenya's Integrated Disease Surveillance and Response (IDSR), the WHO R&D Blueprint, and the Africa CDC priority-pathogen list, and refined through a country-specific feasibility gate that yields the operational surveillance panel and the five antimicrobial-resistance (AMR) marker sets described below; (ii) a 57-site national environmental sampling network spanning Nairobi, Mombasa, Kisumu, Garissa, and Isiolo counties, integrating ILRI's established environmental sites and the WHO Global Polio Eradication Initiative (GPEI) environmental surveillance network; (iii) standardised methods for passive Moore-swab and grab

collection, viral concentration on Nanotrap Magnetic Virus Particles (hydrogel-based affinity capture on paramagnetic beads), nucleic-acid extraction, multi-pathogen screening by TaqMan Array Card (TAC), absolute quantification by digital PCR (dPCR), and broad-spectrum characterisation through metagenomic sequencing; (iv) bacterial culture with whole-genome sequencing for AMR profiling and viral whole-genome sequencing for variant tracking; and (v) a real-time geospatial analytics platform (Geo App) built on ArcGIS Online for visualisation, thresholded alerting, and bilingual (English/Kiswahili) situation reporting. The full cycle from collection to an actionable public health report is executable within 7 calendar days, a turnaround consistent with public health response requirements. The framework is designed for adaptation across low- and middle-income settings and is aligned with emerging global WBE reporting standards.

Keywords: wastewater-based epidemiology, environmental surveillance, Kenya, multi-pathogen surveillance, antimicrobial resistance, poliovirus, metagenomics, digital PCR, TaqMan Array Card, sub-Saharan Africa.

Conflict of Interest Declaration: The authors declare that they have no affiliations with, or involvement in, any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Introduction

The burden of reactive surveillance in Africa

Africa carries a disproportionate burden of infectious diseases and antimicrobial resistance[1–3] and reports more than 100 public health emergencies each year[4], yet the dominant surveillance systems are facility-based, clinician-dependent, and reactive[5], and routinely fail to deliver the early warning required for effective outbreak containment. Most public health practitioners have thus identified that the gap is both quantitative as well as structural; for every confirmed clinical case of a pathogen, such as *Salmonella* Typhi or *Vibrio cholerae*, an estimated 4-100 community cases go undetected, and [6,7] the resulting surveillance architecture is thus blind and silent to the majority of community transmission.

Wastewater-based epidemiology (WBE) has thus been shown to be capable of resolving this blind spot by providing an aggregate pathogen bio-signature shed in a clearly defined population. This principle has been established in environmental science for decades. Pathogens, nucleic acids, and their metabolic by-products excreted by both symptomatic and asymptomatic individuals accumulate in municipal wastewater and can be detected and quantified using molecular methods to provide a time-integrated, population-level sentinel signal[8,9]. WBE is non-invasive, independent of healthcare-seeking behavior, and is thus capable of detecting nascent outbreaks before clinical case counts cross the conventional alert thresholds.

Proof of concept: COVID-19 and the global WBE revolution

The global expansion of WBE has been witnessed during the COVID-19 pandemic. Different studies from Europe, North America, Africa, and Asia have demonstrated that SARS-CoV-2 RNA signals in municipal wastewater consistently precede confirmed clinical cases by 4-14 days, providing actionable early warning[10]. The United States National Wastewater Surveillance System (NWSS), established at the height of the COVID-19 pandemic in 2020, now has more than 1,200 sites serving approximately 140 million people, and has been expanded to include influenza A and B, respiratory syncytial virus (RSV), and mpox [10]. Similarly, the European Centre for Disease Prevention and Control (ECDC) has established a continent-wide wastewater surveillance network, and EU legislation now requires Member States to extend routine wastewater surveillance to additional targets, including antimicrobial resistance[11]. This global experience has established WBE as a tool for modern public health preparedness.

The sub-Saharan Africa gap

Despite global momentum, sub-Saharan Africa remains under-represented in WBE implementation. Although WBE has proved feasible and cost-effective in low- and middle-income settings[12], most African applications remain short-term, research-driven projects rather than routine national surveillance; this pattern documented across recent enteric-virus, tuberculosis, and antimicrobial-resistance reviews and multi-country assessments[13–17]. This inequality is attributed to inadequate laboratory infrastructure for molecular testing, inadequate cold chain logistics for sample transport, limited local adaptation of WBE methods for tropical wastewater matrices with high inhibitor loads, and minimal integration with national public health governance frameworks [18,19].

In Kenya, infectious disease surveillance is primarily reactive. Although the Kenya National Public Health Institute (KNPHI) has established a network of sentinel surveillance sites using the IDSR framework, coverage remains uneven, reporting is delayed, and community-level asymptomatic transmission of certain enteric pathogens and tuberculosis is still being missed. For instance, Kenya has experienced recurring cholera outbreaks (1,498 cases reported in 2023 in programmatic KNPHI data), a persistent polio eradication challenge, endemic *Mycobacterium tuberculosis* with a case-notification rate of 209 per 100,000[20], and an alarming rise in AMR. East Africa has one of the highest documented rates globally for extended-spectrum beta-lactamase (ESBL)- and carbapenem-resistant Enterobacterales (CRE)[3]

Previous WBE work in Kenya

SARS-CoV-2 environmental surveillance in Kisumu (2021-2022) detected viral RNA in 14.3% of samples and showed that each positive detection was associated with a ninefold increase in the odds of identifying clinical COVID-19 cases 4-7 days later[21]. This was followed by a retrospective wastewater study in Nairobi, which similarly confirmed Delta as the dominant SARS-CoV-2 variant circulating between April and August 2021, providing molecular-epidemiological evidence unavailable from clinical testing alone[22]. Previously, metagenomic surveillance in Kibera, Nairobi's largest informal settlement, also demonstrated the feasibility of broad-spectrum pathogen and AMR detection in raw sewage using

shotgun sequencing[23]. None of the three studies, however, constituted a prospective, validated, multi-pathogen national surveillance framework. Critical gaps from the three studies include non-harmonized methods, lack of systematic pathogen prioritization, national sampling-network design, WBE quality-assurance framework, and real-time data integration.

Scope and innovation of this protocol

In this study, we established an end-to-end validated national WBE framework for Kenya. Its innovations are:

- i. Transparent, seven-criterion pathogen prioritization matrix operating over a 64-pathogen candidate universe, drawn from IDSR, WHO R&D Blueprint, Africa CDC, and the WHO Bacterial Priority Pathogens List (BPPL) sources. The matrix produces an auditable composite score and a country-adjustable feasibility gate; therefore, the final surveillance panel is defensible on both epidemiological and operational grounds and reproducible in any LMIC.
- ii. Integrated 57-site national sampling network, combining ILRI's established environmental sites with the WHO GPEI polio-accredited network, maximizes catchment coverage while leveraging existing site permissions, logistics, and community engagement.
- iii. The multi-platform laboratory pipeline was validated for Kenyan wastewater matrices with high inhibitor loads: TAC for high-throughput screening; QIAcuity dPCR for absolute quantification; Oxford Nanopore and MGI DNBSEQ for real-time metagenomics and outbreak investigation; and Illumina DNA Prep whole-genome sequencing with high-resolution bacterial and viral characterization.
- iv. A quality assurance framework anchored by validated internal process controls (bacteriophage MS2), defined acceptance criteria, and inhibitor-management strategies tailored for Kenya wastewater.
- v. The AI-integrated, real-time geospatial analytics platform (Geo App) automates data integration, generates threshold-based outbreak alerts, and produces bilingual (English/Kiswahili) policy-ready situation reports.
- vi. The reflex pathway integrates WBE signals with facility-based diagnostics so that a wastewater signal triggers targeted clinical testing creating a direct link between population-level surveillance and individual patient care.
- vii. A Kenya National Minimum Data Set (MDS) for WBE reporting enforces cross-county comparability, integrates with the electronic IDSR and DHIS2 platforms, and contributes to global WBE repositories.

Overview of the Protocol

The protocol we propose comprises six major phases, from study design to public health reporting (Fig. 1). Phase 1 consists of a pathogen prioritization exercise that identifies the pathogen list from Kenya's IDSR disease and condition lists. We then proceed to select and validate sites for the selected pathogens in Phase 2. Phase 3 covers field sample collection (passive 24-hour Moore swabs or grabs), packaging, cold-chain transport, and laboratory reception. In Phase 4, we describe the laboratory analytical pipeline in which a portion of each specimen was aliquoted for AMR culture, while the second portion was

subjected to viral concentration on Nanotrap Magnetic Virus Particles, RNA extraction (QIAamp Viral RNA Mini Kit), multi-pathogen screening (TaqMan Array Card), absolute quantification (QIAcuity dPCR), and outbreak investigation using Oxford Nanopore rapid barcoding (broad surveillance) or Illumina DNA Prep (high-accuracy bacterial and viral WGS). In Phase 5, we cover data management, statistical analysis, and AI-assisted analytics. Finally, Phase 6 describes the Geo App platform for real-time visualization, expression, and alert generation. End-to-end turnaround in a routine surveillance setup is anticipated to be between 5-7 calendar days from the time of sample collection.

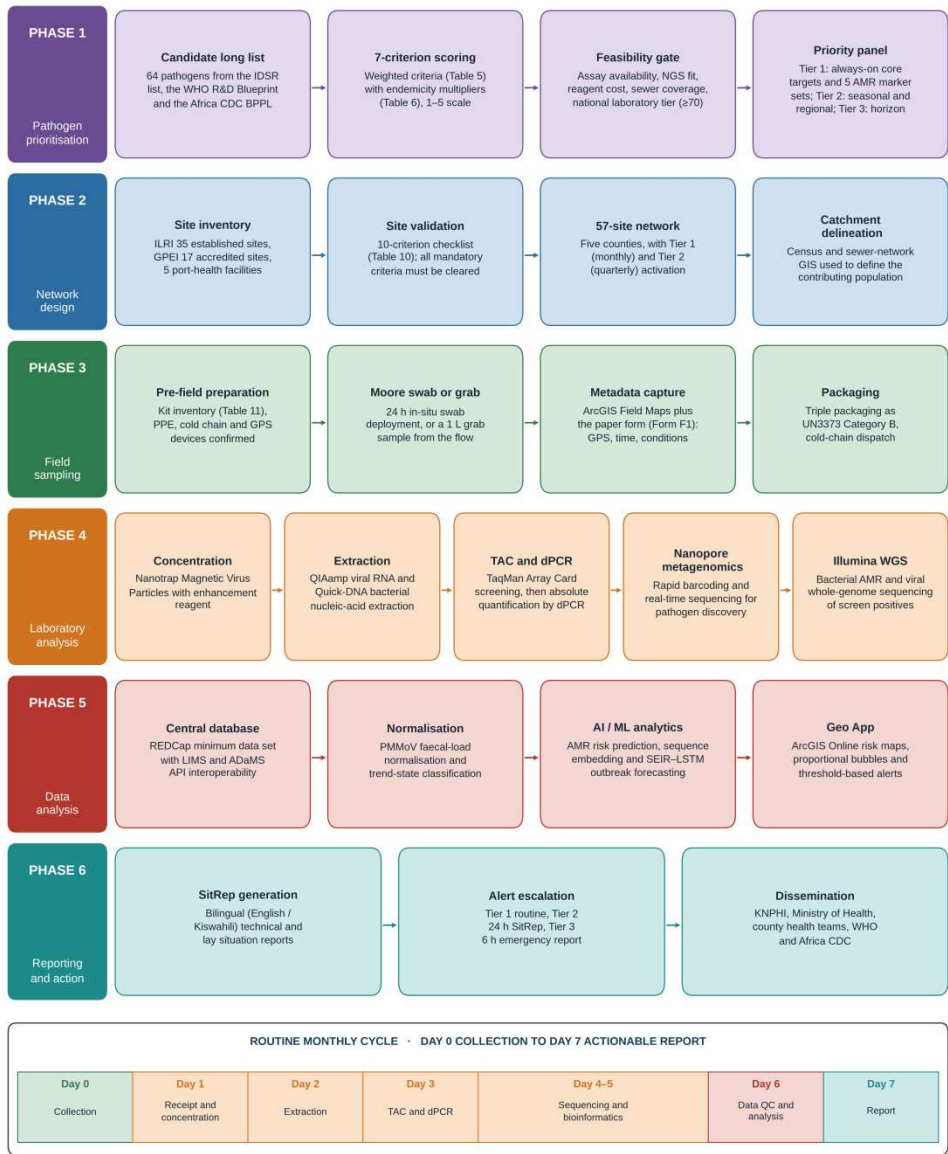


Figure 1:End-to-end workflow of the Kenya National Multi-Pathogen Wastewater-Based Epidemiology (WBE) framework. The schematic illustrates the six-phase process from pathogen prioritisation and network design through field sampling, laboratory analysis, data analysis, and reporting. Turnaround from Day 0 (collection) to Day 7 (report) under routine surveillance conditions

Expertise Needed to Implement This Protocol

The implementation requires a multidisciplinary team. **Table 1** summarizes the core competencies. A single field cycle can be run by a coordinator, plus two field officers per county, and centralized laboratory processing is a 5-6 person team spanning molecular biology, bacteriology, and bioinformatics.

Table 1: Personnel competency requirements for implementing the Kenya National WBE protocol

Role (and function)	Minimum qualification	Key competencies required	Personnel count
Field Sampling Coordinator — lead environmental sample collection (water, air, soil)	MSc in Environmental Science or related field	Field logistics management, aseptic sampling techniques, team leadership, GPS/GIS proficiency	1
Molecular Laboratory Analyst — perform DNA/RNA extraction and qPCR	BSc in Molecular Biology or Biochemistry	Precision pipetting, PCR/qPCR protocols, DNA/RNA quantification, quality-control procedures	3
Bacteriologist — culturing, identification and characterization of bacterial strains	PhD or MSc in Microbiology	Anaerobic culture, biochemical testing, antibiotic susceptibility testing, microbial microscopy	2
Sequencing Specialist — operation and maintenance of next-generation sequencing (NGS) platforms	MSc in Genomics or Genetics	Library preparation (Illumina/Nanopore), sequencing instrument operation (e.g. MiSeq), troubleshooting, sequence data management	1
Bioinformatician — analysis of large-scale genomic data sets	PhD or MSc in Bioinformatics or Computational Biology	Programming (Python, R), assembly algorithms, phylogenetic analysis, metagenomic workflow development	1
Data Manager — management, storage and security of research data	BSc in Computer Science, Data Science, or Information Management	Database management (SQL), data quality assurance, data-sharing protocols (e.g. FAIR), cloud storage administration	1
Epidemiologist — analysis of disease patterns and risk factors related to identified pathogens	PhD in Epidemiology or Public Health	Statistical modelling (SAS/STATA), outbreak investigation, data visualisation, spatiotemporal analysis	1
Biosafety Officer — oversight of laboratory safety protocols and regulatory compliance	MSc or Certified Biosafety Professional (e.g. RBP)	Biosafety Level 2/3 (BSL-2/3) compliance, risk assessment, training, waste-management protocols	1

Limitations

- i. WBE can underestimate community pathogen burdens where open defecation prevalence is high (notably in Garissa and Isiolo); therefore, we interpreted the results alongside sanitation-access data.
- ii. Estimating the contributing catchment population is inherently uncertain in informal settlements, so we will apply multiple normalizers (creatinine, PMMoV, sewer-network trace) and report the resulting range.
- iii. In hot climates, RNA degradation between voiding and collection lowers the apparent signal; therefore, strict collection-to-testing cold-chain logistics will be maintained.
- iv. Clinical case data may be unlikely to match the demographic composition of the wastewater catchment exactly; therefore, cross-referencing WBE signals with case notifications may carry residual confounding.
- v. Certain vector-borne pathogens may be unsuitable for WBE (low shedding into municipal sewage) and are better tracked by entomological or clinical systems.

- vi. Industrial wastewater carries variable PCR inhibitor loads, and as such, site-specific inhibitor profiling is required before routine deployment.
- vii. The full analytical stack requires substantial capital investment; therefore, a resource-constrained deployment can start with concentration, extraction, and RT-qPCR only, adding sequencing capacity once the routine pipeline is established.
- viii. The network is a sentinel design built on existing, largely sewered, environmental-surveillance infrastructure and does not, by itself, sample the majority of the population who rely on non-sewered sanitation; national-level inference will therefore be restricted to the surveillance domain of the contributing catchments, with stratified reporting and phased expansion as the pathway to broader national coverage (see Phase 2).

Materials

Equipment

To implement the protocol, there is a need to acquire laboratory and field equipment for sample processing, molecular analysis, sequencing, and data generation across operational sites, as summarized in **Table 2**.

Table 2: Equipment required for the Kenya national WBE protocol

Equipment	Model / specification	Application	Location
Biosafety cabinet	Thermo Scientific 1300 Series A2	Open sample processing	NGSML, NMRL, ILRI, KEMRI DLSP
Microcentrifuge	Eppendorf 5425	Extraction, library prep	All labs
Benchtop centrifuge	Thermo Fisher Sorvall Legend XTR + TAC buckets	TaqMan Array Card runs	NGSML, KEMRI DLSP
Magnetic separator (15 mL)	Invitrogen DynaMag-15	Nanotrap concentration	NGSML, ILRI
Magnetic separator (1.5 mL)	Invitrogen DynaMag-2	Nanotrap elution, bead cleanup	NGSML, ILRI, KEMRI DLSP
Thermal cycler	Applied Biosystems ProFlex 96-Well	cDNA synthesis, BLT PCR	All labs
Real-time PCR system	Applied Biosystems QuantStudio 7 Flex	TAC qPCR	NGSML, KEMRI DLSP
Digital PCR system	Qiagen QIAcuity One (5-plex)	dPCR quantification	NGSML
Nanopore sequencer	Oxford Nanopore MinION Mk1C or GridION X5	Metagenomic sequencing	ILRI, NGSML, KEMRI POLIO LAB
Illumina sequencer	Illumina MiSeq or NextSeq 550	Bacterial WGS	ILRI, NGSML, KEMRI DLSP
Spectrophotometer	NanoDrop 2000c or Nanophotometer P360	Nucleic-acid quality assessment	All labs
Fluorometer	Qubit 4	Nucleic-acid quantification	All labs
Fragment analyser	Agilent 4150 TapeStation or 2100 Bioanalyzer	Library size verification	ILRI, NGSML, KEMRI DLSP
Bead beater / homogeniser	FastPrep-24 5G	Bacterial cell lysis	NMRL, KEMRI DLSP
Vortex mixer	Scientific Industries Vortex-Genie 2	Sample homogenisation	All labs
Plate shaker	Eppendorf ThermoMixer C	Illumina library prep mixing	NGSML, ILRI
GPS device	Garmin eTrex 32x	Site geolocation	Field teams
Temperature logger	Testo 174H	Cold-chain monitoring	Field teams

Equipment	Model / specification	Application	Location
Insulated transport box	Pelican 1510	Sample cold-chain transport	Field teams

Reagents

In addition to the equipment, a number of reagents and consumables are required, as listed in **Table 3**. Other laboratories interested in implementing the framework can choose substitute reagents, provided they verify the performance against the specified in-house validation criteria (see Reagent Setup and Supplementary Document S1).

Table 3: Reagents and consumables required for the Kenya National WBE protocol

Reagent / consumable	Supplier	Catalogue no.	Storage	Critical notes
Nanotrap® Microbiome Particles	Ceres Nanosciences	44202	4 °C	Vortex well before use. Do not freeze.
Nanotrap® Enhancement Reagent	Ceres Nanosciences	10112	4 °C	Use as directed with Nanotrap particles.
QIAamp Viral RNA Mini Kit	Qiagen	52904	Room temperature (15–25 °C)	Add ethanol to Buffer AW1 and AW2 before use. Ensure a clean workstation to prevent RNase contamination.
Carrier RNA (poly A)	Qiagen	1017647	–20 °C	Aliquot upon receipt. Avoid multiple freeze–thaw cycles.
MS2 Bacteriophage	ATCC®	15597-B1™	–80 °C (long term); 4 °C (working aliquots)	Used as an internal control for extraction efficiency. Monitor for cross-contamination.
Ethanol (molecular biology grade, 96–100%)	Sigma-Aldrich	16606002	Room temperature	Highly flammable. Store in an approved cabinet. Keep bottles tightly sealed to prevent evaporation.
PBS, pH 7.4 (1×)	Thermo Fisher Scientific	10010031	Room temperature	Standard buffering agent for washing and dilution.
Quick-DNA™ Fecal/Soil Microbe Miniprep Kit	Zymo Research	D6010	Room temperature	Requires the addition of ethanol to the Wash Buffer before use.
Beta-mercaptoethanol (β-ME)	Bio-Rad	444203	4 °C	Toxic and strong odor. Use under a fume hood.
TaqMan® Array Card	Thermo Fisher Scientific	Custom (e.g. 4444901)	–20 °C	Keep protected from light. Ensure cards reach room temperature before loading.
TaqMan® Fast Virus 1-Step Master Mix	Thermo Fisher Scientific	4444434	–20 °C	Protect from light. Store in small aliquots. Use within 1 h after removal from freezer.
cDNA Reverse Transcription Kit with RNase inhibitor	Thermo Fisher Scientific	4374967	–20 °C	Keep enzymes on ice during use. Avoid multiple freeze–thaw cycles.
QIAcuity One-Step Viral RT-PCR Kit	Qiagen	1123145	–20 °C	Keep enzymes on ice during use. Protect the reaction mix from light.
Nanoplates, 96-well (QIAcuity)	Qiagen	250021	Room temperature	Do not touch the bottom of the wells. Ensure plates are flat during sealing and cycling.
Custom primers and TaqMan probes	Integrated DNA Technologies	Custom	–20 °C	Store at high concentration. Aliquot working stocks. Protect probes from light.
Rapid Barcoding Kit (SQK-RBK004)	Oxford Nanopore Technologies	SQK-RBK004	–20 °C	Follow manufacturer protocols exactly. Do not vortex enzymes.
AMPure XP Beads	Beckman Coulter	10136224	4 °C	Equilibrate beads to room temperature (20–25 °C) for at least 30 min before use.
MinION Flow Cell (R9.4.1)	Oxford Nanopore Technologies	FLO-MIN160D	4 °C	Keep in a sealed pouch until use. Avoid thermal pouch.

Reagent / consumable	Supplier	Catalogue no.	Storage	Critical notes
Qubit™ 1X dsDNA HS Assay Kit	Thermo Fisher Scientific	Q33230	Room temperature (assay reagent at 4 °C)	Use only thin-wall polypropylene tubes. Handle standards carefully.
Illumina DNA Prep Kit (24-Sample)	Illumina	20060059	-20 °C	Store indexes separately. Follow purification steps precisely.
IDT for Illumina Indexes (e.g. UD Indexes)	Integrated DNA Technologies for Illumina	20027213 for set A,	-20 °C	Single use. Avoid cross-contamination.
Microseal® 'B' Adhesive Seals	Bio-Rad	0030127781	Room temperature	Apply firm, even pressure to seal. Ensure optical clarity.
Moore Swabs, Sterile	[to be confirmed]	Made in the lab from gauze	Room temperature	Handle using sterile technique, For environmental sampling.
HDPE Bottles 1 L	Thermo Fisher	2104-0032	Room temperature	Use sterile bottles for sample collection.
Nitrile Gloves (powder-free)	[to be confirmed]	9.006 375 for small, 9.006 376 for medium, 9.006 377 for large and 9.006 378 for X- large	Room temperature	Wear during all lab procedures. Change frequently.
N95 Respirator Masks	[to be confirmed]	1860	Room temperature	For respiratory protection when handling potentially infectious samples.
Thermo-logger (USB temperature data logger)	[to be confirmed]	17953982	Room temperature	Place with temperature-sensitive reagents during transport and sampling.

Reagent Setup

We describe the setup applied to the Kenya National Public Health Institute National Genomics and Molecular Laboratory (NGSML). We use manufacturer instructions and kit inserts to guide our set-up: QIAamp Viral RNA Mini Handbook (QIAGEN, 2019); Quick-DNA Fungal/Bacterial Miniprep Handbook (Zymo Research, 2023); Nanotrap Magnetic Virus Particles Instructions for Use (Ceres Nanosciences, 2023); QIAcuity OneStep Advanced Probe RT-PCR Kit Handbook (QIAGEN, 2022); and the TaqMan Array Card Getting Started Guide (Thermo Fisher Scientific, 2021) are some of the guides we use to develop our standard operating procedures. From the various options provided in the instruction for use (IFU), NGSML chose the following options:

Carrier RNA (QIAGEN, catalogue #1017647) was reconstituted per the QIAamp handbook by adding 310 µL of Buffer AVE to the 310 µg vial, yielding a 1 µg/µL stock. We divided this stock into 10 µL single-use aliquots and stored them at -20 °C. On the day of extraction, we added carrier RNA to Buffer AVL at 10 µL per 1 mL of AVL. Buffers AW1 and AW2 receive 96-100% molecular-grade ethanol before first use, per the handbook labels on each bottle, and are then stable at room temperature for the shelf life stated by the manufacturer.

For Zymo Genomic Lysis Buffer, we add beta-mercaptoethanol to a final concentration of 0.5% (v/v) immediately before use, as specified in the Quick-DNA handbook. Working 80% ethanol was prepared

fresh from 100% molecular-grade ethanol and nuclease-free water on the day of use, with at least 10 mL prepared per sample to cover both washing steps.

Nanotrap Magnetic Virus Particles are supplied as a hydrogel-based affinity-capture reagent covalently linked to paramagnetic beads, and they carry a dye-modified affinity bait that binds low-abundance viral particles in complex matrices, such as wastewater[24]. The particles were equilibrated to room temperature for 15 min and vortexed for 10s immediately before dispensing so that the bead suspension was homogeneous. Nanotrap Enhancement Reagent 2 is added at the manufacturer-specified volume to promote binding of enveloped viruses; the bound viral fraction is then recovered on a magnetic separation rack, and free wastewater matrix (including PCR inhibitors) is aspirated and discarded.

Bacteriophage MS2 (ATCC 15597-B1) was used as an internal process control by diluting the working stock to 10^6 PFU/mL and spiking 10 μ L (10^4 PFU) into each sample tube at the start of the concentration step. Unused working aliquots were frozen at -80 C, and thawed aliquots were never re-frozen. The full step-by-step procedures for each reagent's current IFU are recorded in Supplementary Document S1.

Primer and probe reference table

In Table 4, we list the primer and probe sequences used for the molecular detection of each priority pathogen and AMR marker, along with the target gene, expected amplicon size, and primary literature reference. We validated the selected primer sets in wastewater; however, each new deployment must verify the limits of detection, linearity, and specificity against spiked local wastewater using both purified target nucleic acids and, where available, whole inactivated pathogens before routine reporting.

Table 4:Primer and probe sets used in the Kenya national WBE panel. Each row lists the target pathogen, gene target, amplicon length, and primary reference. Local in-house verification (LOD, linearity, specificity) is mandatory before routine reporting

Pathogen	Target gene	Forward primer (5'→3')	Reverse primer (5'→3')	Probe (5'→3')	Size (bp)	Reference
SARS-CoV-2 (N gene)	N gene	GACCCCAAAATCAGCGAAAT	TCTGGTTACTGCCAGTTGAATCTG	FAM-ACCCCGCATTACGTTTGGTGGACC-BHQ1	72	CDC 2019-nCoV_N1 assay
SARS-CoV-2 (N2)	N gene (N2)	TTACAAACATTGGCCGCAAA	GCGCGACATTCCGAAGAA	FAM-ACAATTTGCCCCAGCGCTTCAG-BHQ1	67	CDC 2019-nCoV_N2 assay
Influenza A	M gene	GACCRATCCTGTCACCTCTGAC	AGGGCATTYTGACAAAKCGTCTA	FAM-TGCAGTCTCTGCTCACTGGGCACG-BHQ1	100	CDC Flu assay
Influenza B	M gene	TCCTCAACTCACTCTTCGAGCG	CGGTGCTCTTGACCAAAATTGG	FAM-CCAATTCGAGCAGCTGAACTGCGG-BHQ1	110	CDC Flu assay
Measles virus	N gene	CACCATCCAAGATGTTCTTGCC	GATAGTTGGAGGCGGGAAGG	FAM-TCTCCGAGAAATGGTCGATGTGCC-BHQ1	130	Hummel et al., 2009
Poliovirus (pan)	5'UTR	CAAGCACTTCTGTTTCCCC	CAAGTAACACCCGTGTTTCTCC	FAM-TTGTACCATAAGCAGCCA-BHQ1	100	WHO GPEI standard assay
Vibrio cholerae	ctxA	CTGAATACAGAGGCTTATAAAG	GCAATAGCCCCAGCGTTTA	FAM-CAGCTTCTCAGTCAGTTGTAGTG-BHQ1	148	Bhatt et al., 2007
Salmonella Typhi	staA	TCCAAGCAACCAAGATAAG	AGTCTTCCGTAATCGTCTTCC	FAM-CAATGATTCGCTGGCAGTACAGC-BHQ1	120	Buckle et al., 2012
Shigella spp.	ipaH	GTTCTTGACCGCCTTTCGGAT	GCCGGTAAAACCGTCTCCGT	FAM-CCCGGATGGTATGGGCGTTCCGG-BHQ1	147	Garaicoa Pazmino et al., 2013
M. tuberculosis (MTB)	IS6110	GTCATGGCCGCGCTTTCGT	CAAAGCACCGCTAAAGCTCCT	FAM-ATGACCTTCGCTCCACCAAGGT-BHQ1	75	Piersimoni & Scarpato, 2003
M. tuberculosis (RIF-R)	rpoB (RRDR)	TCGCGGCCGATCAAGGAGT	ACGTGCTCATCGACAGCGGA	FAM-TGCACGTCGACGACAACAGACATCG-BHQ1	90	Ravigione et al., 2012
NDM-1 (CRE)	blaNDM-1	GGTTTGCGGATCTGGTTTTTC	CGGAATGGCTTCATCACCATC	FAM-AGTTTGGCTGCCTTGTGTCAT-BHQ1	88	Poirel et al., 2011
KPC (CRE)	blaKPC	TCGCTAAACTCGAACAGG	TACTGCCCGTTGACGCCC	FAM-TGCTCACTGGCTGTGGTGCGGAG-BHQ1	120	Bratu et al., 2005
CTX-M (ESBL)	blaCTX-M	GGGTGACGAAACGACTCTTCC	CGCATTGACGTTACTCTACTCT	FAM-AACGTGGTGGCATCGGATTATTCG-BHQ1	540	Pitout et al., 2004
MS2 (internal control)	mat protein	ATAGGTGGGAGAGGAGTTT	AATCTCAATGAGCGAACCC	HEX-AGCGATCTTGTCCAGATCTCCATCGAC-BHQ1	99	Haramoto et al., 2018

Procedure

Phase 1 - Pathogen prioritization framework

In the initial phase of the study design, prioritization will take 2-4 weeks, and this procedure needs to be refreshed annually or in response to emerging threats and evidence. We based our WBE pathogen panel Programme on a seven-criterion matrix developed, as indicated in the table below.

The seven-criterion prioritisation matrix

We used pre-published thresholds to score on a scale of 1-5 for each candidate pathogen against seven weighted criteria. The criteria and their default weights for a Kenya-stage Pilot Programme (Pathway A) are listed in Table 5. The weights we have indicated are editable and can be re-tuned when a country transitions to an “expanding” Programme (Pathway B), which increases the weight of epidemic potential and information gain relative to the raw disease burden.

Table 5: Weighted criteria applied to every candidate pathogen. Default weights are shown for the Kenya Pilot (Pathway A) and for the 'Expanding' (Pathway B) programme stages. The remaining weight - Information Gain (0.04 / 0.12) and Operational Readiness (0.15 / 0.04) - is applied at the panel stage to guard against including targets whose WBE signal is already well captured by clinical surveillance or that a national laboratory cannot yet assay

Criterion	What it measures	Weight (Pilot/Expanding)	Example raw indicators
1. Disease burden	Incidence per 100k; case-fatality; U5 mortality contribution.	0.16 / 0.16	IHME GBD 2021 incidence; WHO case data.
2. Epidemic potential	Outbreak frequency, R0, PHEIC history.	0.14 / 0.16	WHO DON; R0 from outbreak literature.
3. WES suitability	Shedding site and load, environmental stability, WASH sensitivity.	0.18 / 0.10	Published shedding studies; persistence data.
4. Clinical surveillance gap	IDSR strength, laboratory capacity, health-system reach.	0.10 / 0.13	WHO IHR SPAR C05 / C04 (2023 round).
5. Actionability	Availability of vaccine, treatment, or public-health response.	0.10 / 0.10	WHO PIP / EPI programme registers.
6. AMR / treatment-failure risk	WHO BPPL priority class, resistance emergence trend.	0.05 / 0.10	WHO BPPL 2024; GLASS reports.
7. Vulnerable-population equity	Under-5 mortality; sanitation gap; fragility.	0.08 / 0.09	IGME U5MR; JMP; FSI.

Endemicity-aware objective

Because the value of a WBE signal depends on whether a pathogen is endemic, sporadic, emerging, or absent in the target country, each pathogen row in our prioritization matrix carries an endemicity classification and a corresponding scoring objective, such as trend intensity monitoring, early detection sensitivity, horizon scanning, or watch only, respectively. We created endemicity multipliers (Table 6) to reweight the seven criteria so that endemic pathogens are scored primarily on burden and equity,

sporadic and emerging pathogens are scored primarily on epidemic potential and information gain, and pathogens not currently present in the country are down-weighted across the board.

Table 6:Endemicity-mode multipliers applied on top of the base weights in Table 5. Multipliers are supplied by the calculator workbook and can be edited to reflect national epidemiological context.

Criterion	Endemic	Sporadic	Emerging	Not Present
Burden	1.4	0.6	0.3	0.1
Epidemic potential	0.7	1.5	1.4	0.2
WES suitability	1.0	1.2	1.2	0.5
Clinical surveillance gap	1.1	1.2	1.0	0.3
Actionability	1.3	1.0	0.7	0.2
AMR risk	1.2	0.9	0.6	0.2
Equity	1.2	0.9	0.6	0.2
Information gain (panel-level)	1.0	1.2	1.3	0.3
Operational Readiness (Panel-Level)	1.0	1.0	1.1	0.5

From Raw Indicator to 1-5 Score

We derived each criterion's 1-5 score from raw indicators using explicit thresholds so that the process was reproducible. The key thresholds we used for the Kenya calculations are summarized in Table 7 and are drawn from IHME GBD 2021, WHO IHR SPAR (2023 round), WHO/UNICEF JMP (2024), Fragile States Index (2023), WHO DON archive (2021-2026), and published pathogen-shedding studies.

Table 7:Threshold table used to convert raw indicators into 1-5 sub-scores. The full 21-row cut-off list, including the sub-scores for individual indicators such as sewer-connected population and NGS capacity, is provided in the accompanying calculator workbook(Supplementary File S1).

Indicator	1	2	3	4	5
Incidence per 100,000	<10	>=10	>=50	>=200	>=500
Case-fatality rate (%)	<1	>=1	>=3	>=10	>=30
Outbreaks last 10 years	0	>=1	>=3	>=5	>=8
R0	<1.2	>=1.2	>=2	>=3	>=5
Shedding sum (stool+urine+resp)	<2	>=2	>=3	>=4	>=6
Environmental stability (days)	<3	>=3	>=7	>=14	>=30
WASH sensitivity (0-1)	<0.2	>=0.2	>=0.4	>=0.6	>=0.8
Country surveillance strength (inverse)	5	4	3	2	1
Vaccine + treatment availability	None	Supportive	Tx only	Vaccine only	Both
WHO BPPL tier	NA	Low	Medium	High	Critical
Under-5 mortality (per 1000)	<15	>=15	>=30	>=50	>=80

Basic sanitation coverage (%)	>=80	<80	<60	<40	<25
Fragility index	<50	>=50	>=70	>=85	>=100

Feasibility gate

A pathogen only enters the operational panel if it passes an independent feasibility gate a 0-100 score composed of assay availability (30%), NGS fit (25%), reagent cost band (15%), sewer catchment coverage (15%), and national laboratory tier (15%). Only pathogens scoring ≥ 65 on composite priority and ≥ 70 on feasibility are classified as “Do now” in the current program stage. Pathogens with high priority but insufficient feasibility are categorized as “low-cost add-on” (multiplex onto an existing panel when capacity permits) or 'Defer' (revisit at the next program review). This detailed work is provided in the prioritization Excel workbook.

Application to Kenya

We then ran the prioritization calculator with country = Kenya and program stage = Pilot (Pathway A) and pulled the country parameters from the sources documented in the workbook, which are reproduced in Table 8.

Table 8:Kenya country parameters that drive the pathogen prioritisation calculation. Values are drawn from the workbook's 'Sources' tab and the Kenya country reference row (extraction date 2026-07-30).

Kenya country reference parameter	Value	Source
Population (millions)	57.5	UN DESA WPP / World Bank WDI (2025 est.)
Under-5 mortality (per 1000)	38.8	UN IGME (2024)
Basic sanitation coverage (%)	40.9	WHO/UNICEF JMP (2024)
Basic drinking-water coverage (%)	65.6	WHO/UNICEF JMP (2024)
Health index (blended HAQ + UHC SCI)	45.2	IHME GBD 2019; WHO GHO (2023)
IHR SPAR C05 – Surveillance	4	WHO e-SPAR 2023 round
IHR SPAR C04 - Laboratory capacity	4	WHO e-SPAR 2023 round
WHO-notified outbreak events, 5-year	3	WHO DON archive 2021-2026
Fragile States Index	87.8	Fund for Peace FSI (2023)
Sewer-connected population (%)	15	WHO/UNICEF JMP + UN-Habitat (2024)
Established polio ES sites	12	GPEI (2025)
National NGS capacity (0-3)	3	Africa CDC Pathogen Genomics Initiative (2024)
National lab-network tier (0-3)	3	e-SPAR C04 (2023) + Africa CDC mapping

The Kenya WBE Priority Panel

We obtained a ranked composite score in Figs. 2 and 3 and the priority panel in Table 9 after running the calculator against the 64-pathogen candidate list. Our selected pathogen set is dominated by high-shedding, high-burden targets that already have validated WES assays and clear public health action pathways: poliovirus (WPV1, cVDPV2), SARS-CoV-2, influenza A (including H5N1), RSV, adenovirus (enteric and respiratory), norovirus, rotavirus, *Vibrio cholerae*, *Salmonella* Typhi, *Shigella* spp., *Mycobacterium tuberculosis*, and the ESBL/CRE AMR marker set (blaCTX-M, blaTEM/SHV, blaNDM, blaKPC, and blaOXA-48). Tier 2 ('Seasonal / Regional') adds hepatitis A/E, enterovirus, *Cryptosporidium*, *Giardia*, *Schistosoma*, Ebola (event-driven), Lassa (event-driven), and mpox. Tier 3 was reserved for horizon scanning of currently absent pathogens (Nipah and Guinea worms).

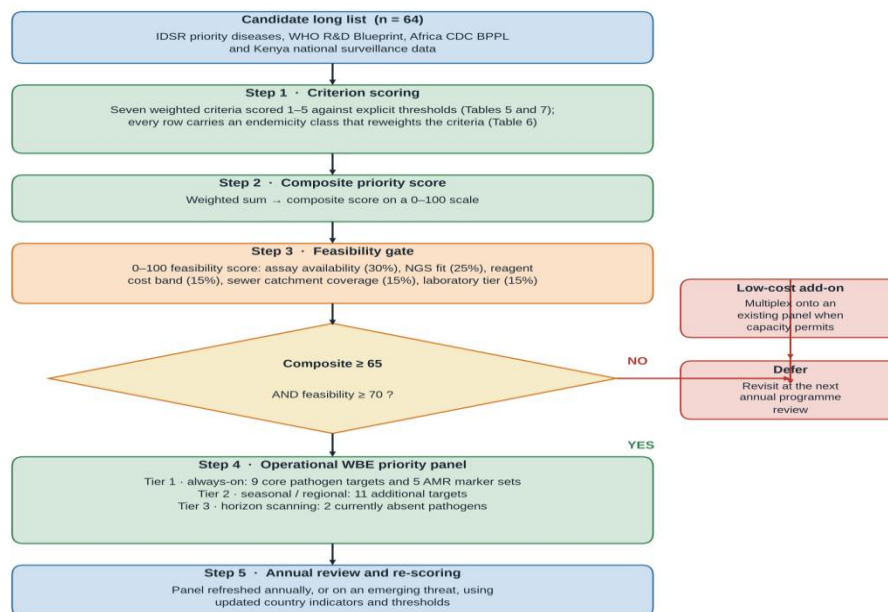


Figure 2: Five-step pathogen prioritisation workflow, from the IDSR long list through the seven-criterion score and feasibility gate to the final WBE priority panel.

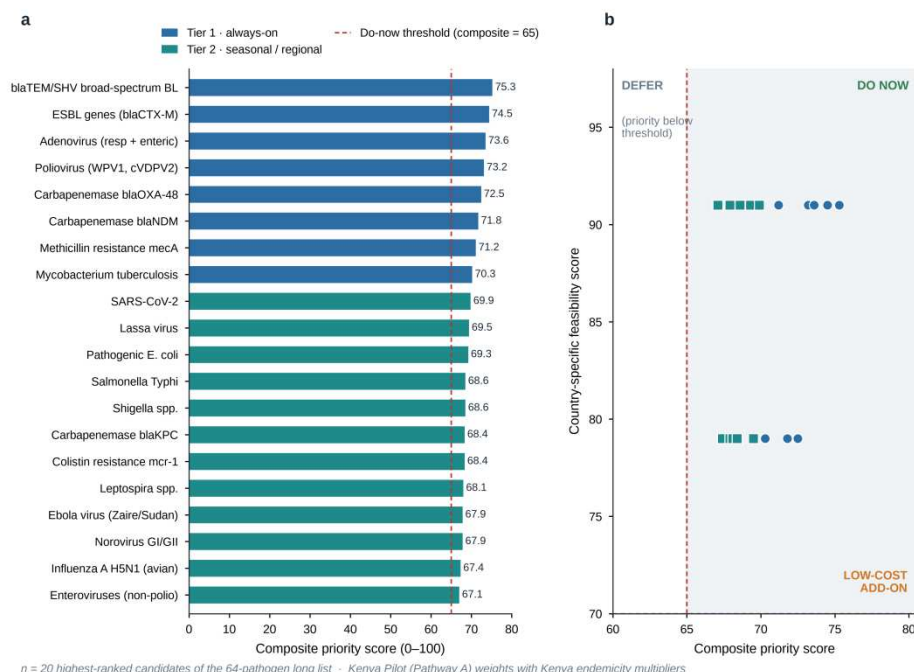


Figure 3:Kenya pathogen prioritisation, showing performance against the five short-form scoring criteria used in the summary tile (C1 public-health significance, C2 analytical feasibility, C3 shedding dynamics, C4 operational feasibility, C5 public-health actionability)

Table 9:Top-ranked 20 pathogens for the Kenya Pilot (Pathway A) programme, drawn from the 64-pathogen calculator output. Composite is on 0-100 using Pilot-stage weights and Kenya endemicity multipliers; feasibility is the 0-100 country-specific score. 'Do now' requires composite >=65 AND feasibility >=70. The complete ranked list of all 64 candidate pathogens is provided as Supplementary Table S1

Pathogen / AMR marker	Composite (0-100)	Tier	Feasibility	Quadrant
blaTEM/SHV broad-spectrum BL	75.3	T1 Always-on	91	Do now
ESBL genes (blaCTX-M)	74.5	T1 Always-on	91	Do now
Adenovirus (resp + enteric)	73.6	T1 Always-on	91	Do now
Poliovirus (WPV1, cVDPV2)	73.2	T1 Always-on	91	Do now
Carbapenemase: blaOXA-48	72.5	T1 Always-on	79	Do now
Carbapenemase: blaNDM	71.8	T1 Always-on	79	Do now
Methicillin resistance: mecA (MRSA)	71.2	T1 Always-on	91	Do now
Mycobacterium tuberculosis	70.3	T1 Always-on	79	Do now
SARS-CoV-2	69.9	T2 Seasonal	91	Do now
Lassa virus	69.5	T2 Seasonal	79	Do now
Pathogenic E. coli (ETEC/EPEC/EHEC)	69.3	T2 Seasonal	91	Do now
Salmonella Typhi	68.6	T2 Seasonal	91	Do now
Shigella spp.	68.6	T2 Seasonal	91	Do now

Carbapenemase: blaKPC	68.4	T2 Seasonal	79	Do now
Colistin resistance: mcr-1	68.4	T2 Seasonal	79	Do now
Leptospira spp.	68.1	T2 Seasonal	79	Do now
Ebola virus (Zaire/Sudan)	67.9	T2 Seasonal	79	Do now
Norovirus GI/GII	67.9	T2 Seasonal	91	Do now
Influenza A H5N1 (avian)	67.4	T2 Seasonal	79	Do now
Enteroviruses (non-polio)	67.1	T2 Seasonal	91	Do now

Phase 2 - Study Site Selection and Network Design

Based on our work plan, the surveillance network will be established over 4-6 weeks and updated iteratively as the data yield accumulates. Our WBE network combines two pre-existing environmental surveillance infrastructures: ILRI's 35 established environmental sampling sites, WHO GPEI's 17 accredited environmental poliovirus surveillance sites, and five port health facilities. This yields a 57-site national network across the five counties, as indicated in Fig. 4, maximizing coverage while capitalizing on existing sampling logistics, site permissions, and community engagement relationships.

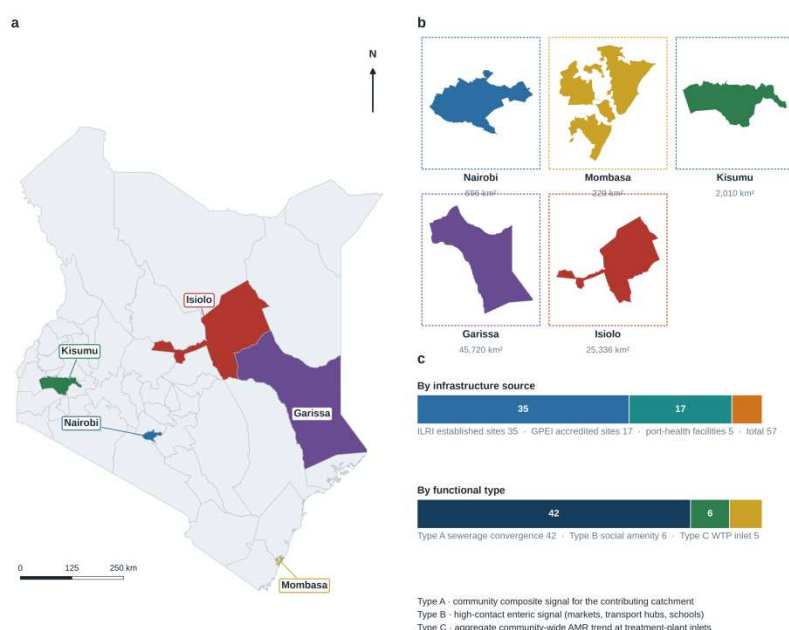


Figure 4:Kenya national WBE surveillance network - 57 sites classified into three functional categories: Type A sewerage convergence points (n = 42), Type B social-amenity sites (n = 6), Type C wastewater-treatment-plant inlets (n = 5). Each type serves a distinct epidemiological objective (community-level composite, high-contact enteric signal, and aggregate community-wide AMR trend, respectively).

Representativeness and inferential scope. The network is a nationally distributed sentinel system, not a random probability sample of the Kenyan population: only about 15% of Kenya's population is connected to sewer networks, and open defecation remains common in parts of the country. The 57 sites therefore

characterise pathogen circulation in the sewered and institutionally connected catchments they serve (predominantly urban and peri-urban populations), and national-level statements will be limited to this surveillance domain rather than extrapolated to the whole population. Three design measures mitigate the resulting selection bias: national summaries will be stratified by county and site category (sewer-convergence, social-amenity, and treatment-plant-inlet sites) so that no national statement rests on a single site class; WBE signals will be interpreted alongside facility-based surveillance data (eIDSR, DHIS2) that cover non-sewered populations; and the network will expand in phases to additional counties and, as validated methods are established, to non-sewered settings (open drains and peri-latrines) in counties with high open-defecation prevalence.

Site validation procedure

In the workplan, we scored every candidate site against the ten-criterion validation checklist indicated in **Table 10**. A site will have to clear all mandatory criteria before Tier 1 (monthly) activation; one or more conditional findings will drop it to Tier 2 (quarterly) surveillance, pending remediation; rejection of any mandatory criterion (accessibility, biosafety, authorization) blocks activation.

Table 10: Site validation checklist. Only sites that clear all mandatory items on the first pass are activated as Tier 1 routine-monthly sites

Validation criterion	Assessment method	Threshold for acceptance	Classification
Physical accessibility	Site visit and inspection	Safe and easy access for personnel	Mandatory
Moore swab fixtures	Verification of installation	Swabs installed in all designated locations	Mandatory
Flow continuity	Continuous monitoring data review	Uninterrupted flow during sampling periods	Conditional
Human faecal origin indicators	Microbial source tracking (MST)	Confirmed presence of human-specific faecal markers	Conditional
Catchment population	Data analysis and mapping	Accurate population estimation of the service area	Mandatory
GPS coordinates	Precision GPS measurements	Accurate to within the specified variance range	Conditional
Biosafety risk	Risk assessment protocol	All potential risks identified and mitigation measures implemented	Mandatory
County health authorization	Documentation review	Valid and official authorization permit obtained	Mandatory
Community sensitisation	Verification of outreach efforts	Evidence of community awareness and engagement activities	Mandatory
Site photographs	Visual documentation check	Clear, dated photographs for each site location	Conditional

Catchment population delineation

Once a site has met all the criteria, we will perform population normalization, which is critical for converting copies per litre into meaningful per-capita metrics such as estimated infections per 100,000. We obtained sewer-network GIS shapefiles from the Nairobi City Water and Sewerage Company, the Coast Water Services Board, and the Lake Victoria South Water Services Board and traced the upstream network from every sampling point in ArcGIS Network Analyst (with QGIS used as a cross-check). Once the catchment polygon is defined, we overlay it on the Kenya National Bureau of Statistics (KNBS) 2019

census enumeration areas to obtain the resident population. We will then validate each catchment estimate against three orthogonal fecal-load indicators: daily creatinine concentration (expected 1.0-1.5 g per person per day), ammonia nitrogen concentration, and pepper mild mottle virus (PMMoV) concentration, and record the estimate together with the methodology and the uncertainty range in the site characterisation form. This will be refreshed annually against the KNBS inter-census projections.

Phase 3 - Field Sample Collection

Each monthly collection cycle will take 2-3 hours per site, including travel, collection, packaging, and documentation. The sampling workflow is illustrated in **Fig. 5**.

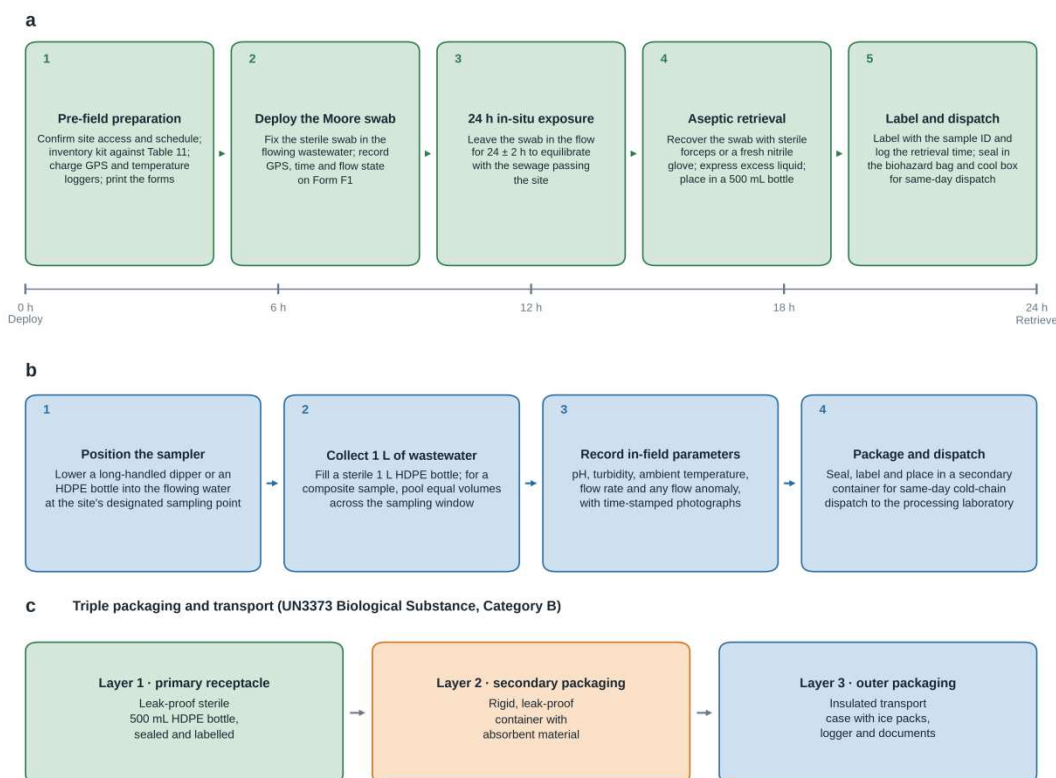


Figure 5: Complete field sampling workflow, from pre-field preparation through Moore-swab deployment or grab collection to packaging and cold-chain dispatch

Pre-field preparation

The field coordinator will confirm the following at least 24 h before the scheduled sampling day: (i) the sampling schedule and site access; (ii) courier bookings for Day 1 pickup; (iii) that PPE stocks (double supply of nitrile gloves, N95 respirators, coveralls, rubber boots, and face shields) are adequate; (iv) the field kit is inventoried against **Table 11**; and (v) the temperature loggers are charged and zeroed and the digital data-capture devices (GPS units or smartphones running ArcGIS Field Maps) are fully charged. Sample-collection forms will then be printed, and when commercial Moore swabs are not being used, Moore swabs will be prepared and sterilized in advance.

Table 11:Field sampling kit inventory checklist - per field team, per sampling day

Item	Quantity required	Check (□)
Sterile Moore swabs	1 per site + 3 spares	
Sterile 1 L HDPE bottles	2 per site + 3 spares	
Sterile 500 mL HDPE bottles	1 per site + 3 spares	
Sterile ziplock biohazard bags	2 per site + 5 spares	
80% ethanol spray bottle	2	
Nitrile gloves	4 pairs/person + spares	
N95 masks	2/person	
Disposable coveralls	2/person	
Safety goggles	1/person	
Rubber boots	1 pair/person	
Utility hook	1/site + 2 spares	
Long-handled dipper	2	
Waterproof marker	3	
Pre-printed labels	2 sets/site	
Sample collection forms	1/site + 5 blanks	
GPS device	1/team	
Temperature logger	1/box	
Insulated transport box	1 per 6–8 samples	
Secondary containment bags	1/site + 3 spares	
Biohazard waste bags	3	
Sharps container	1	
Emergency spill kit	1	
Chain of custody forms	1 set/consignment	
Portable field notebook	1	

Moore-swab deployment

We will treat every open wastewater access point as a BSL-2 hazard, and as such, we will ensure that the following are catered for: full PPE (double nitrile gloves, N95 respirator, face shield, coveralls, rubber boots) is mandatory, personnel do not work alone, and a second team member will be stationed upwind to provide safety cover and, if needed, summon emergency assistance. The Moore swab method depends on passive adsorption of microorganisms from flowing wastewater over 12-24 hours; Therefore, we first confirm that the access point is actively flowing and that the flow depth is sufficient to submerge the swab. The full deployment SOP, including retrieval and immediate insertion into the pre-labelled Whirl-Pak bag, with the recording of retrieval time, is shown in Supplementary Document S2.

Grab sample collection

Where the Moore swab is not applicable, we will implement a grab sample and capture a discrete instantaneous volume. We fixed the collection window at 06:00-10:00 local time (the morning peak of

domestic discharge) across every site and logged the exact collection minute in the metadata record. The complete grab-sample SOP is provided in the Supplementary Document S3.

Sample metadata recording

Every collected sample will have a standardized metadata set as summarized in **Fig. 6**, recorded in full on the paper collection form (Supplementary Form F1) and entered concurrently into ArcGIS Field Maps for auto-synchronization with the central database.



Figure 6: Metadata elements required for each WBE sample record, spanning sample identity, environmental context, chain of custody, laboratory procedures, and results linkage

Sample packaging and transportation

Once the sample has been collected, it is shipped as UN3373, Biological Substance Category B. The personnel will ensure triple packaging (primary leak-proof container inside a secondary sealed pouch inside a rigid outer shipping carton with absorbent material sufficient to absorb the entire liquid content in the case of a primary breach). Failure to comply with these standards resulted in carrier rejection and sample loss. The sample-tracking form is also accompanied by every shipment. We have provided a detailed sample transportation SOP in the Supplementary Document S4.

Sample reception, triage, and distribution

All samples will be received at the NGSML and distributed to the ILRI, the National Microbiology Reference Laboratory (NMRL), and KEMRI for downstream processing. Sample acceptance and rejection criteria (visible degradation, temperature excursion beyond 4 ± 2 C, chain-of-custody discontinuity, or missing metadata) will be enforced before any downstream work. All rejected samples are documented and, where possible, re-collected at the next scheduled cycle. The full triage and distribution of SOP is provided in Supplementary Document S5.

Phase 4 - Laboratory Analytical Pipeline

In **Fig. 7** below, we provide a summarized laboratory workflow overview.

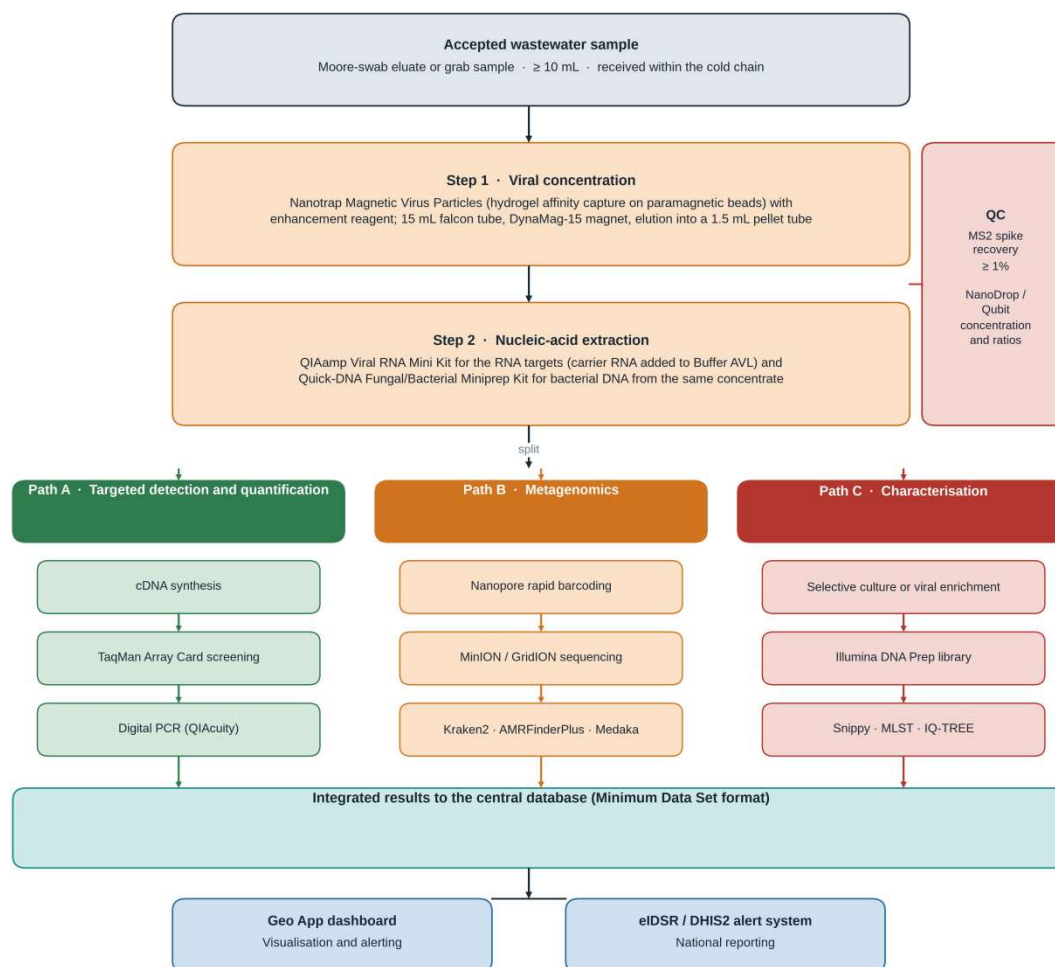


Figure 7: Laboratory analytical pipeline. Sample intake feeds two parallel tracks - Path A (targeted detection by TAC and dPCR) and Path B (metagenomics by Oxford Nanopore or MGI DNBSEQ) - with Path C (bacterial and viral whole-genome sequencing on Illumina) providing high-resolution characterisation of Path A positives.

Viral concentration on Nanotrap Magnetic Virus Particles

All viral concentrations will be performed in a Biosafety Level 2+ (BSL-2+) standards laboratory using a Class II Type A2 biosafety cabinet, with all open-tube steps confined to the cabinet. Using Ceres Nanosciences IFU (2023), we will equilibrate the hydrogel-coated Nanotrap Magnetic Virus Particles to room temperature and vortex, and then add Nanotrap Enhancement Reagent 2 to promote the binding of enveloped viruses. The bead–virus complex was then incubated in a magnetic rack, and the supernatant was discarded. The beads were then eluted in a small volume for downstream extraction. This magnetic-capture step efficiently concentrated low-abundance viruses while removing most of the wastewater matrix and PCR inhibitors. We have provided the full step-by-step SOP in the Supplementary Document S6.

Viral RNA extraction - QIAamp Viral RNA Mini Kit

We then extracted batches of ten samples using the QIAamp Viral RNA Mini Kit per the QIAGEN handbook. Under the chaotropic conditions of the kit, viral RNA is adsorbed onto the silica spin-column membrane, while impurities are removed by sequential washing steps (AW1 and AW2). Purified RNA was eluted in Buffer AVE and immediately transferred on ice for downstream RT-qPCR and dPCR benches. All open-tube RNA work will be confined to an RNase-free zone, and each batch will have a negative control (NC) and positive extraction control (spiked MS2). We have provided a detailed batch-processing SOP in Supplementary Document S7.

Bacterial DNA extraction - Quick-DNA Fungal/Bacterial Miniprep Kit

Apart from the sample being shared for concentration, part of the wastewater sample was shared with the Microbiology Laboratory for culture work. Bacterial isolates recovered from the culture (*Vibrio cholerae*, *Salmonella Typhi*, *Shigella* spp., and CRE isolates) were collected and stored. After 24-48 hours of culture activation, we will run ten isolates in parallel per the Zymo Research handbook: mechanical and chaotropic lysis, silica column binding, sequential wash, and low-salt elution. Purified DNA was collected and used for downstream qPCR and Illumina library preparation without further purification. We have provided the detailed SOP in the Supplementary Document S7.

Note: We will start processing for *Mycobacterium tuberculosis* (Tier 3, **exploratory target**) only in the KNPBI BSL-3 facility once the renovation is complete. Liquid MGIT cultures were heat-inactivated at 85 °C for 60 min before being transferred out of the BSL-3 containment, per the KNPBI biosafety policy. The detailed BSL-3 extraction SOP is provided in the Supplementary Document S8.

Multi-pathogen detection - TaqMan Array Card (TAC)

We loaded the RNA extract onto the TAC card and performed PCR in a Thermo Fisher Q7 for approximately 4 h; total per-batch turnaround: approximately 7 h (8 samples per card). The TAC platform will allow the simultaneous qPCR of up to 48 pathogen-specific assays on a single microfluidic 384-well card, and it is our primary high-throughput screening tool before confirmatory dPCR quantification. It uses hydrolysis-probe cleavage for detection during amplification. We documented the TAC assay layout, run parameters, and interpretation rules in Supplementary Document S9.

Absolute quantification - QIAcuity RT-dPCR

We will follow every TAC-positive priority pathogen using the QIAcuity Digital PCR System with the QIAcuity OneStep Advanced Probe RT-PCR Kit per the QIAGEN handbook. In the one-step format, viral RNA is reverse-transcribed into cDNA and amplified inside thousands of nanoliter partitions, allowing absolute copy quantification without a standard curve. Compartmentalized chemistry will also provide better tolerance to residual inhibitors, a well-documented advantage for wastewater matrices. The full run parameters, per-target thresholds, and QC acceptance rules are indicated in Supplementary Document S9.

Rationale: Although TAC provides a qualitative detection status, dPCR delivers absolute quantification (copies per μL , no external calibration), which makes cross-site, cross-time, and per-litre-of-wastewater comparisons defensible.

Metagenomics - Oxford Nanopore Rapid Barcoding and Illumina NextSeq 1000

TAC-positive samples and a monthly pool of TAC-negative samples will also be subjected to metagenomics as an ongoing quality assurance for the TAC panel. We used the Oxford Nanopore Rapid Barcoding Kit, which uses transposase to fragment DNA and simultaneously attach barcode adapters, allowing multiple samples to be pooled and run on a single R9/R10 flow cell (MinION or GridION). We base all the reads in real time. The flow cell time is extended for samples still gaining unique targets and is terminated when the target coverage curve plateaus. In parallel, the Illumina Nextseq 1000 will be used for short-read platforms, where high-accuracy short reads are required for AMR gene detection or lineage typing. We have detailed the platform-specific SOPs in Supplementary Documents S10 (Nanopore) and S11 (MGI DNBSEQ).

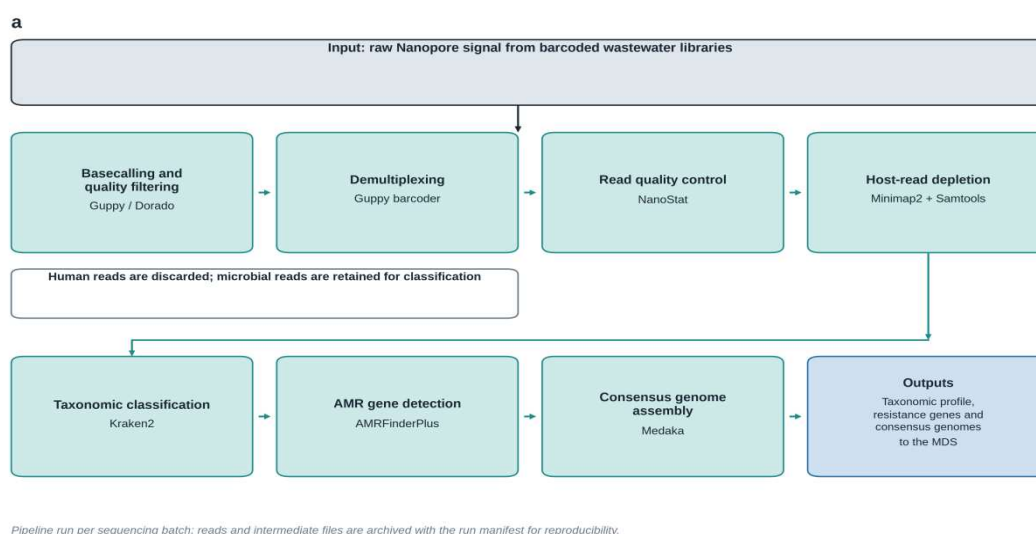


Figure 8: Bioinformatics pipeline for Nanopore metagenomics: basecalling and quality filtering (Guppy/Dorado), demultiplexing, quality control (NanoStat), host-read depletion (Minimap2 + Samtools), taxonomic classification (Kraken2), AMR-gene detection (AMRFinderPlus), and consensus genome assembly (Medaka)

Bacterial and viral whole-genome sequencing - Illumina DNA Prep and MGI DNBSEQ

We selected priority viral samples from Path A and all bacterial isolates recovered from culture and performed whole-genome sequencing (WGS) using the Illumina DNA Prep IFU, enzymatic tagmentation of fragment genomic DNA, and simultaneous attachment of adapter sequences, which were then amplified and dual-indexed for multiplexing. The generated libraries were then loaded onto a NextSeq 1000 flow cell and sequenced. This is schematically presented in Figures 8,9 and 10. MGI DNBSEQ workflow was used as an alternative platform for the same libraries. We have provided the full sequence of SOPs in Supplementary Documents S12 (Illumina) and S11 (MGI).

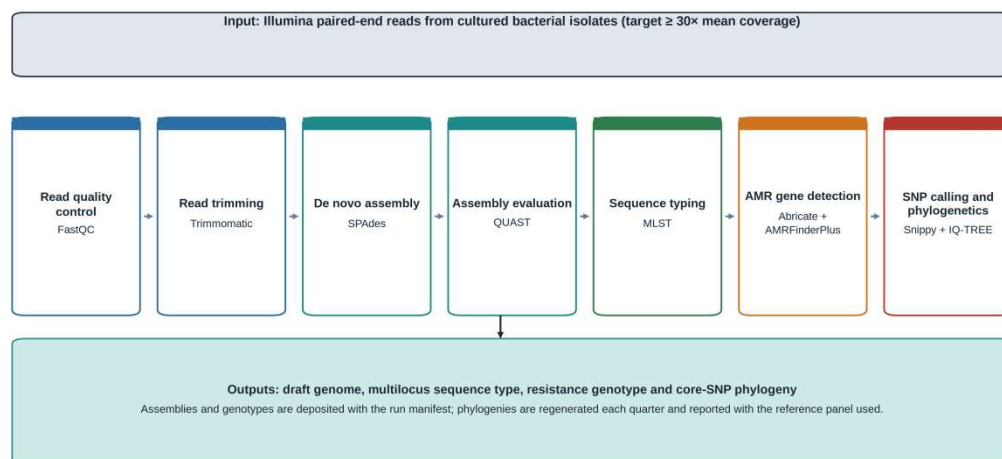


Figure 9: Bacterial WGS bioinformatics pipeline: FastQC and Trimmomatic (quality control), SPAdes (de novo assembly), QUAST (assembly evaluation), MLST typing, Abricate and AMRFinderPlus (AMR gene detection), and Snippy + IQ-TREE (SNP calling and phylogenetics)

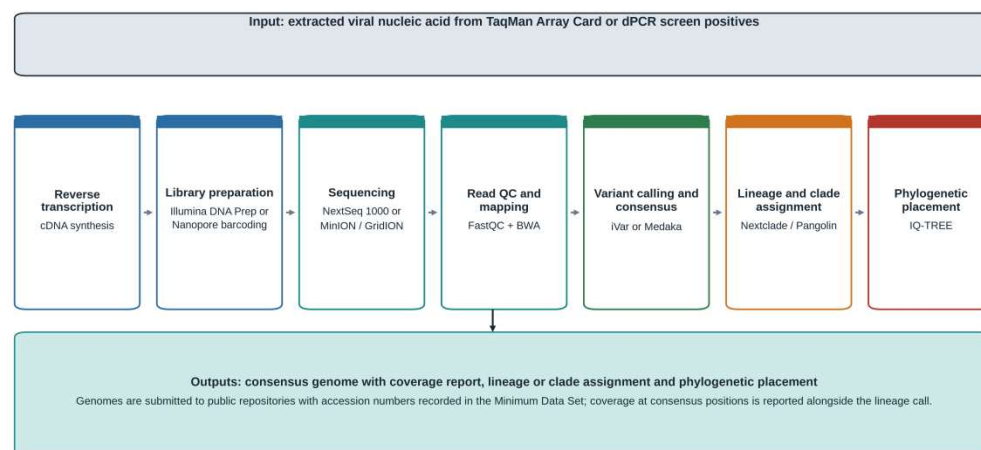


Figure 10: Viral WGS pipeline for lineage and variant identification.

Phase 5 - Data Management, Statistical Analysis, and AI-Assisted Analytics

Central database and Minimum Data Set (MDS)

All WBE data will be entered into a centralized REDCap database hosted on the KNPHI servers. The database is structured around the Kenya National WBE Minimum Data Set (MDS), which was developed through in-country consultation on the essential core elements that every reporting entity must capture. We will use REDCap as it enforces mandatory-field validation, range checks on quantitative fields, removes duplicate detection, and facilitates real-time completeness scoring at data entry. The MDS is designed to be able to interoperate with the Laboratory Information Management System (LIMS) and the All Data Module systems (ADaMS) platform via a dedicated API, so national dashboards are populated automatically without manual re-entry. This is schematically presented in Figure 11.

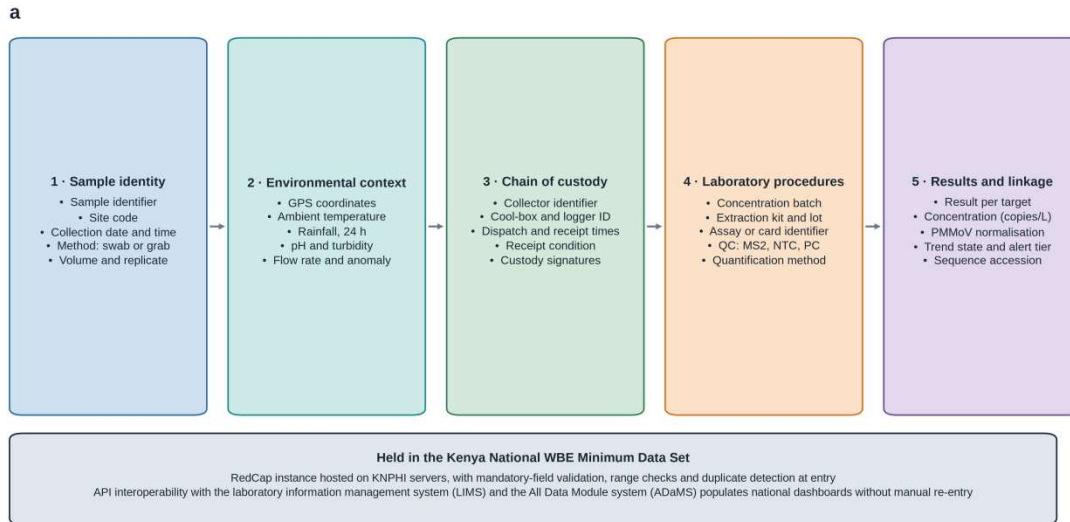


Figure 11:Kenya National WBE Minimum Data Set - the five core blocks captured for every sample

Statistical analysis framework

Sample size and detection power considerations

The network and sampling-cycle design is justified by the detection performance it must deliver rather than by conventional prevalence-estimation sample-size formulas, because routine surveillance is powered for early signal detection rather than for population prevalence estimation. For each Tier 1 site, we will quantify the minimum community prevalence of shedders detectable with 80% probability in a monthly cycle, given the site's catchment population, measured wastewater flow, published per-capita shedding rates, and the analytical limit of detection (LOD) of each assay, using the standard dilution relationship (expected concentration = prevalence × shedding rate × catchment population / daily flow). Site-level detection sensitivity will be recomputed whenever measured flow or catchment population changes materially, and sites whose sensitivity is inadequate for their assigned priority pathogens will be flagged for redistribution or supplementary grab sampling at the quarterly review.

At the network level, the five-county, 57-site design provides at least one monthly (Tier 1) site per participating county and at least twelve samples per site per year, which is sufficient to establish seasonal baselines, estimate within-site variance, and detect sustained shifts in detection rate, while remaining within the laboratory's per-cycle capacity. Because sensitivity additionally depends on shedding heterogeneity and environmental decay, we treat the first twelve months of operation as a calibration phase, after which observed variances, detection rates, and missingness will inform an annual simulation-based power assessment (10,000 resamples per pathogen-site combination) to confirm that the network remains adequately powered for its priority targets and to adjust sample numbers, site tiers, or panel composition where it does not. The full calculation specifications are provided in the pre-specified analysis plan (Supplementary Document S13).

Pathogen detection rate

For each pathogen p at site s in time window t , the detection rate is determined as follows:

$$\text{Detection rate } (p, s, t) = (\text{positive samples} / \text{total samples tested}) \times 100 \%$$

We display the detection rates as per-pathogen and per-county time series. Group comparisons will use chi-square tests, while Fisher's exact test will be substituted when the expected cell count is <5 . Finally, we will use logistic regression models to examine the drivers of detection, with rainfall, ambient temperature, and population density as covariates.

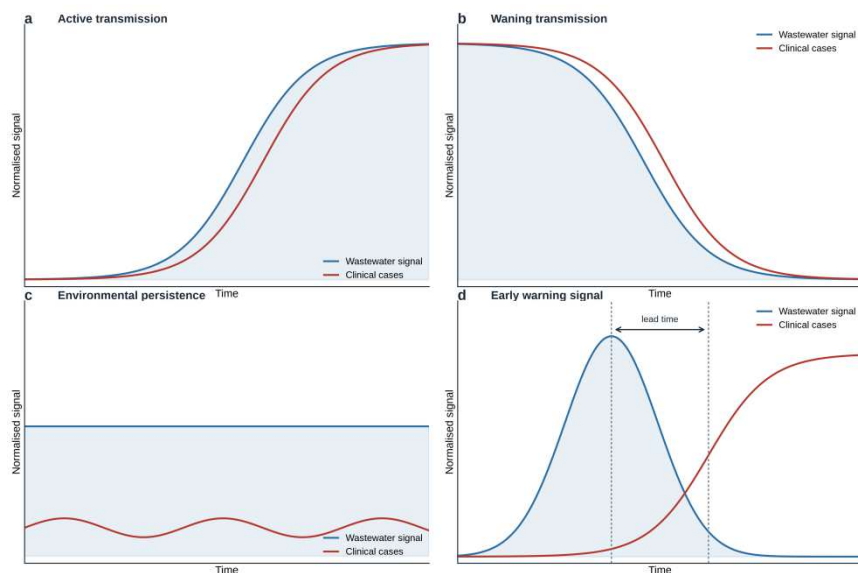
Concentration normalisation

Raw pathogen concentrations will be normalised for dilution and catchment size using PMMoV as a fecal load reference:

$$\text{Normalised concentration} = (\text{copies/L pathogen} \times \text{catchment population}) / (\text{copies/L PMMoV})$$

Infection dynamics inference

Finally, we infer community-level infection dynamics from normalized concentrations using the four canonical trend states illustrated in **Fig 12**: active transmission, waning transmission, environmental persistence, and early warning signal.



Schematic representation of the four canonical trend states used by the trend-state classifier; the classifier assigns a state to each site at every weekly data refresh from the normalised concentration series and the matched clinical series.

Figure 12: Four canonical wastewater-vs-clinical trend states used to interpret WBE signals in near-real time. The early-warning state (WBE spike leading to clinical cases) is the operationally valuable one for outbreak response

Operationally, each site will be assigned to one of the four states by a pre-specified decision algorithm applied to the normalised concentration series: (i) a signal is preliminarily flagged when the site-normalised concentration exceeds the site's established baseline by a z-score of ≥ 2 , or when the robust (Theil–Sen) slope over the last four sampling points exceeds zero at a one-sided α of 0.05; (ii) the early-warning state is assigned when such a rising WBE signal occurs with flat or falling clinical case counts in the catchment, and it persists until the clinical data catch up or the signal resolves; (iii) the active-transmission state requires a rising WBE signal concordant with rising clinical cases; (iv) the waning-

transmission state requires declining WBE and clinical trends; and (v) environmental persistence is assigned when a detectable, non-rising WBE signal persists for ≥ 4 consecutive sampling cycles without a matching clinical signal. Where clinical comparator data are missing or discordant (for example, a clinical rise with no WBE rise), the site is classified as indeterminate and referred to the monthly technical review. The classifier is implemented as a scheduled script in the Geo App backend; states are recomputed at every weekly data refresh, and each SitRep displays the state, the supporting statistics (z-score, slope, and estimated lead time), and a data-confidence flag for every site.

AMR rate calculation

Aside from the infection dynamic inference, we will compare AMR gene prevalence across time points using Mann-Whitney U tests (robust to the non-normal distributions typical of wastewater concentrations). Longitudinal trends will be modelled with ARIMA time-series models, and Principal Component Analysis (PCA) will be used to visualize co-circulation patterns among pathogens and AMR targets. We will then abstract data from clinical sites and use Spearman correlations to link wastewater AMR gene prevalence with sentinel-site clinical AMR data.

Two statistical properties of these data require explicit handling. First, many genes and targets are compared across sites and time points, so we will control the false-discovery rate across all within-target-family comparisons using the Benjamini–Hochberg procedure ($q < 0.05$), reporting effect sizes with 95% confidence intervals alongside q-values. Second, AMR gene-abundance data are compositional (each sample's gene proportions are constrained by the total signal recovered) and are sensitive to sequencing-depth variation; gene abundances will therefore be normalised and analysed on log-ratio scales, with centered log-ratio transformation for principal component analysis and Bayesian multiplicative replacement of zeros where required (additive log-ratio alternatives assessed as sensitivity analyses). ARIMA models will be fitted to log-transformed series, PCA will be performed on centered log-ratio-transformed profiles, and Spearman correlations with clinical AMR data will be computed on log-ratio coordinates; metagenomic abundances will additionally be expressed per million quality-filtered reads to reduce depth-driven artefacts.

AI and machine-learning integration

All AI and machine-learning components will be executed under a single pre-specified analysis plan (Supplementary Document S13) that is registered before model development begins, to constrain overfitting and optimistic performance estimates in the high-dimensional genomic and AMR data space. For each supervised task, the plan predefines the input feature sets and their derivations; outcome definitions and exclusion criteria; training, validation, and test partitions, including temporal splits in which models are trained on earlier data and evaluated only on later data; hyperparameter search spaces and tuning procedures (nested cross-validation within the training partition only); regularisation and early-stopping rules; performance metrics (AUROC, the area under the precision-recall curve, sensitivity and specificity at pre-set thresholds, and calibration); and benchmark comparisons against simple baselines before any model is adopted operationally.

Multiple-comparison control will be applied across model-by-outcome comparisons using the Benjamini–Hochberg false-discovery-rate procedure. Model behaviour will be interrogated before deployment using permutation importance and SHAP analyses, and all model versions, training data, and evaluation results will be archived with the protocol; deviations from the plan will be documented and reported. Prospective evaluation during the Nairobi pilot will be required before any model informs a routine alert.

AI-based AMR prediction

Using the scikit-learn and PyCaret Python libraries, we trained an ensemble of 15 supervised classifiers (including Random Forest, XGBoost, and Support Vector Machines) on antibiotic susceptibility testing (AST) data and interpreted it against Clinical and Laboratory Standards Institute (CLSI) breakpoints. Model outputs are per-pathogen AMR risk scores that flag unusual resistance trends for early warning. The developed model behavior will be interrogated with SHAP (SHapley Additive exPlanations) plots so that public health analysts can inspect which features are driving any given alert. This is schematically presented in Figure 13

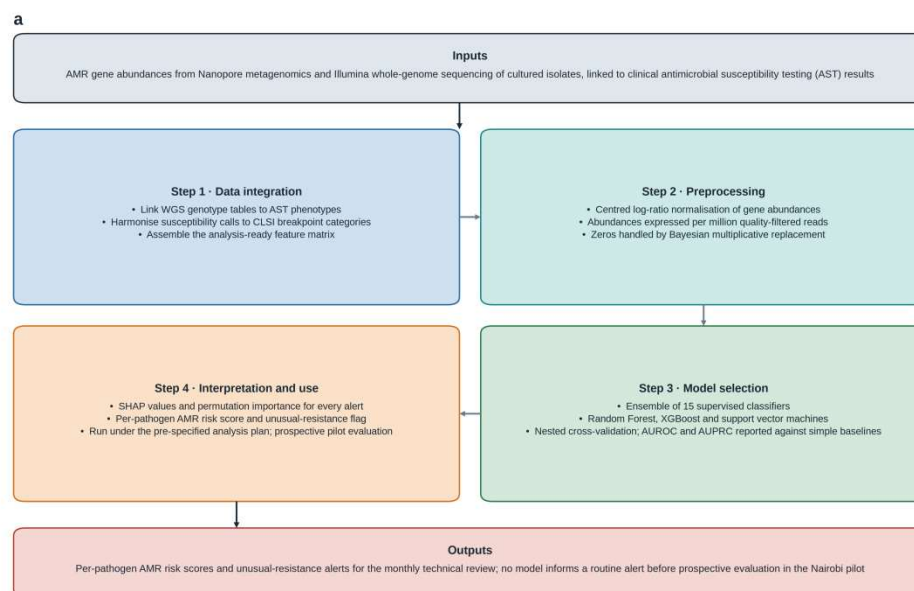


Figure 13:AMR-surveillance machine-learning pipeline, fed by both Nanopore and Illumina WGS data from cultured samples

AI-assisted NGS analytics

We will use DNABERT resources in the International Livestock Research Institute to embed sequences into a high-dimensional vector space for unsupervised phylogenetic clustering to support the early identification of novel mutations and divergent variants that do not cluster tightly with any reference genome and to improve inter-county lineage tracking, accelerate detection of recombinant strains, and accelerate genome classification across both viral and bacterial pathogens.

AI-enhanced disease modelling and triangulation

Hybrid models combining classical Susceptible-Exposed-Infectious-Recovered (SEIR) epidemiological frameworks with Long Short-Term Memory (LSTM) neural networks will be used to forecast outbreak trends and transmission intensity from the WBE signal as shown in **Fig 14**.

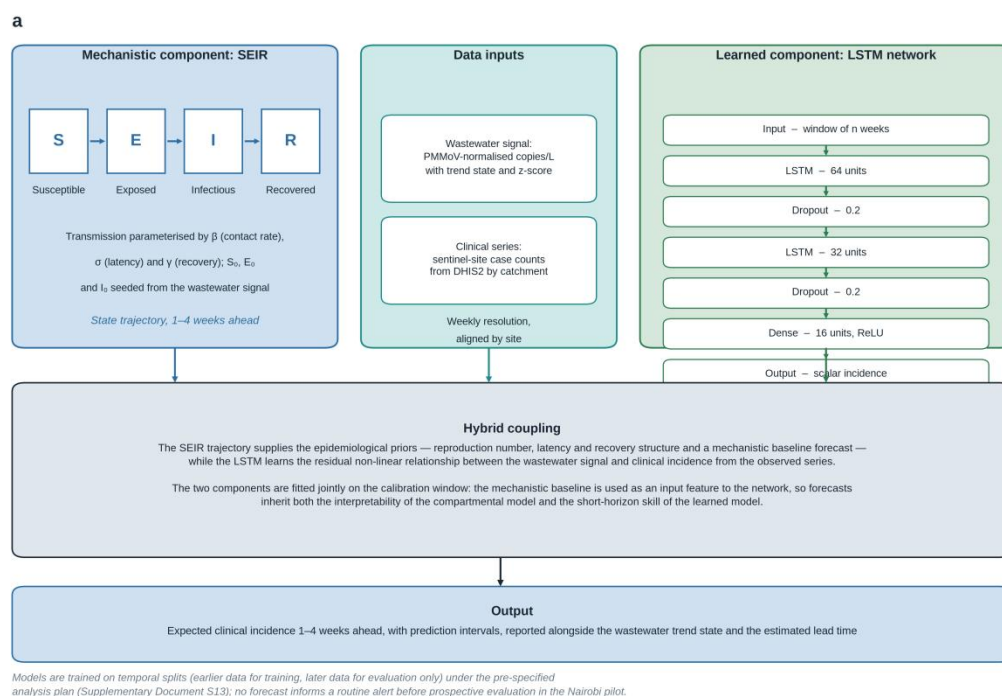


Figure 14:Hybrid SEIR-LSTM forecasting network with LSTM layers, dropout regularisation, dense layer, and scalar output forecasting expected clinical incidence

AI for policy dissemination and decision support

The Geo App will be coupled to an AI-driven Situation Report generator that produces concise, bilingual (English/Kiswahili) briefings for the Director General Kenya National Public Health Institute and public health intelligent officers. Two audience registers are produced in parallel: a technical version for laboratory and epidemiology teams and a lay version for county health officers and public health communication units.

Phase 6 - Geo App: Real-Time Geospatial Analytics and Public-Health Reporting

The Geo App will be built in ArcGIS Online to combine a hosted feature layer and a web map that contains 57 site locations, road networks, administrative boundaries, wastewater-network layers from the participating utilities, and census population data. A GeoForm will surface the data-entry workflow to trained field analysts so that laboratory results are pushed directly to the hosted feature layer with automatic timestamps and geolocation. The App then produces spatiotemporal risk maps (green/amber/red per site), proportional-bubble overlays for pathogen concentration, and per-site and per-county temporal trend lines for easy visualization.

Alert thresholds and escalation protocol

We classified WBE alert signals into three tiers based on pre-specified concentration and trend thresholds, as shown in **Fig15**.

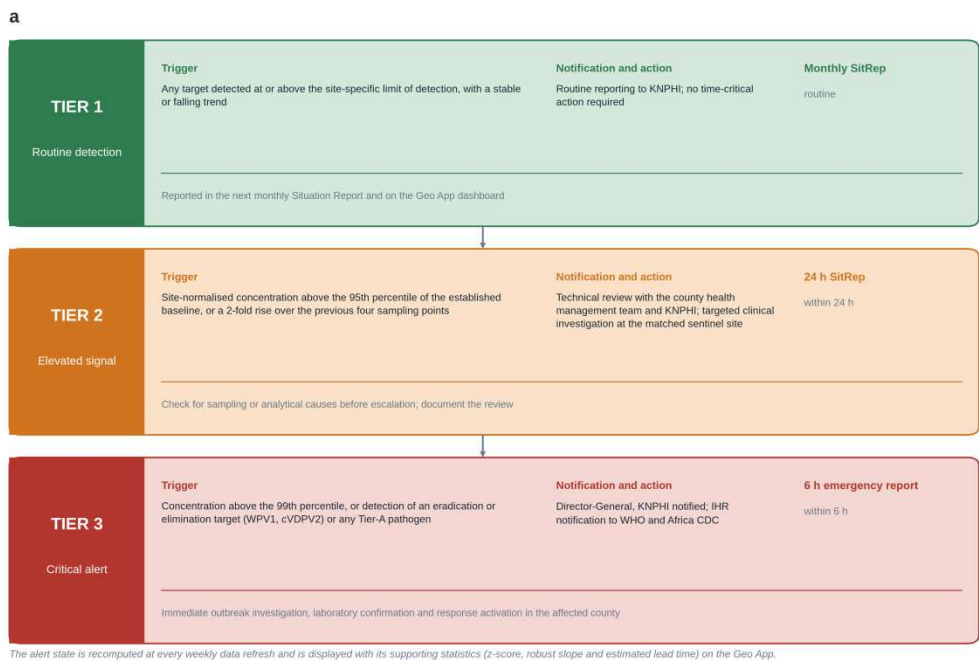


Figure 15:Three-tier alert framework. Tier 1: baseline routine detection reported to KNPHI. Tier 2: >95th-percentile signal, 24-hour SitRep. Tier 3: >99th-percentile signal or Tier-A pathogen detection, 6-hour emergency report and WHO notification

Standardised Situation Report (SitRep) structure

We plan to prepare a monthly SitRep containing the following sections: an executive summary (bilingual) with headline findings, alert signals, and recommendations; a national surveillance overview map covering all 57 sites; pathogen-specific results with detection rates, mean copies/L (with 95% confidence intervals), and comparative trend directions; an AMR summary reporting CRE marker detection frequencies (blaNDM, blaKPC, blaOXA-48), ESBL (blaCTX-M), and MDR-TB (rpoB) prevalence per county with baseline comparisons; a genomic surveillance section highlighting new lineages, variants, and AMR genotypes; a clinical cross-reference section comparing WBE signals with DHIS2 sentinel-site clinical cases, including lead-time estimates; a Tier 2 / Tier 3 alerts log with response actions taken; specific, time-bound recommendations addressed to named implementing entities; and finally a data-quality summary.

Training and capacity building for Geo App use

We will also organize training modules covering data upload, map interpretation, alert thresholds, SitRep use, and eIDSR integration, which are delivered as three-day cohort workshops for KNPHI and county staff. Additionally, we will pilot test the Geo App runs in Nairobi County during the first six months, with a phased roll-out to Mombasa, Kisumu, Garissa, and Isiolo in subsequent quarters conditional on pilot feedback and infrastructure readiness.

Data Reporting and Dissemination

Standardised reporting framework

We organize a reporting framework that is aligned with the emerging best practices in WBE[25,26] and organized into three workflow stages.

- i. Pre-analytical stage: sample ID assignment, GPS coordinates, collection method, weather conditions, and custody chain
- ii. Analytical stage: concentration method, extraction kit, assay identifiers (primers/probes), QC metrics (MS2 recovery, NTC / PC results), quantitative results in copies/L with the calculation method, and sequencing coverage depth.
- iii. Post-analytical stage: normalization method, detection rate estimation, spatial risk maps, and clinical data contextualization.

National Minimum Data Set (MDS)

Our WBE MDS will be developed through (i) stakeholder workshops with county health teams, KNPHI laboratory personnel, and other KNPHI staff; (ii) alignment with the eIDSR data standards; and (iii) design considerations for scalability from sub-national (county) to national reporting, and for adaptability across the East African region.

Dissemination channels

We use a multi-stakeholder architecture, as illustrated in Fig. 16, to share the output of the WBE framework, which maps iterative information flow between real-time surveillance, government oversight, the scientific community, data repositories, and public engagement channels.

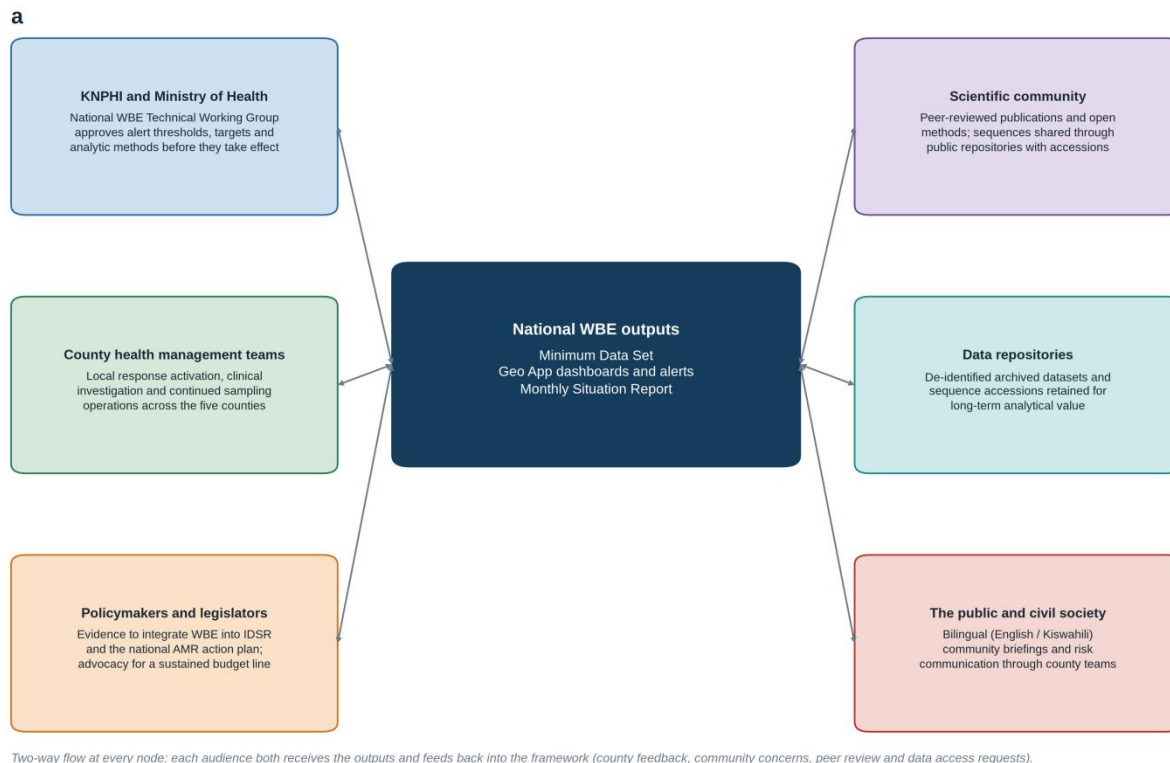


Figure 16: Dissemination architecture linking real-time surveillance to policy action: WBE outputs feed KNPHI/MoH governance; academic and research collaboration for peer-reviewed evidence; secure repositories for long-term analytical value; and policymaker / public- engagement channels that translate evidence into legislative and behavioural response.

Expected Outcomes and Impact

Knowledge outputs

We intend to generate national baseline data on the community prevalence of *SARS-CoV-2*, influenza A/B, poliovirus, measles virus, mpox, *V. cholerae*, *S. Typhi*, *Shigella* spp., *M. tuberculosis*, and AMR markers across five Kenyan counties. The protocol has been optimised for tropical wastewater matrices, making the methods reproducible across sub-Saharan Africa. Additionally, we will deploy a government-owned National Geo App to deliver operational real-time WBE data and systematic quantification of AMR gene prevalence in wastewater to deliver community-wide AMR burden estimates independent of clinical testing access. Genomic epidemiological data will also characterize circulating viral and bacterial lineages far beyond what routine clinical surveillance can deliver.

Anticipated surveillance outcomes

It is expected that this work will demonstrate a documented outbreak-detection lead time aligned with the 7-1-7 detection matrix and the need for using WBE to complement clinical surveillance for priority pathogens across all five counties over the three-year study horizon. This will facilitate hotspot identification through sub-county geographic mapping of high-burden circulation zones, supporting the targeted deployment of clinical resources and vaccination campaigns. Additionally, the work will sustain

poliovirus eradication through continuous environmental surveillance at all 57 sites, complementing GPEI-accredited locations, and extending the geographic coverage of the Kenya National Polio Eradication Programme. In addition, this work will provide AMR stewardship intelligence through community-level tracking of AMR gene prevalence, feeding directly into the Kenya National AMR Action Plan, and county-level stewardship interventions.

Policy and systems-change outcomes

The integration of the WBE into Kenya's IDSR framework will provide an evidence base for adopting the WBE as a mandated complementary surveillance component. By the end of Year 3, capacity building will have empowered health departments in all five participating counties to sustain their local WBE sampling operations and incorporate them into their annual work plans and budgets. The project will deliver a fully documented, costly, and validated framework replicable by other LMIC national public health institutes across sub-Saharan Africa and beyond.

Timing

The end-to-end operational timeline is a three-year implementation cycle running from 2026 to 2029, as indicated in Fig. 17. The activity-level Gantt chart is presented in Supplementary Table S2.

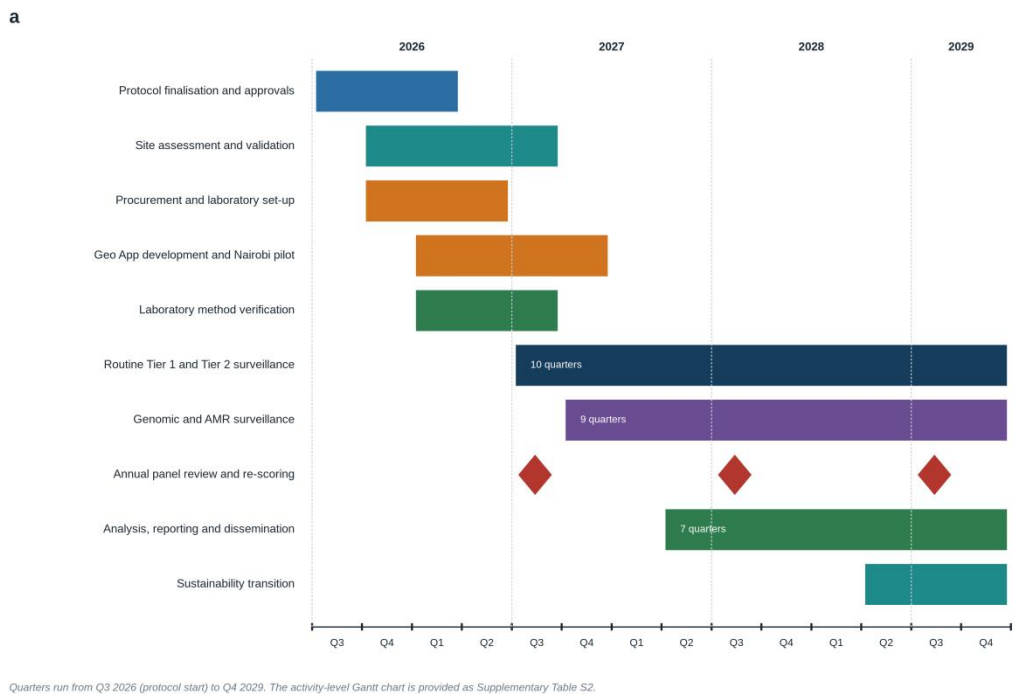


Figure 17:Kenya national WBE study implementation timeline (2026-2029).

Per-cycle timing for routine monthly surveillance

The protocol was designed for rapid, integrated surveillance. As summarized in Fig. 18, the routine monthly cycle takes 5-7 calendar days from field collection to the bio informatics-verified report for routine surveillance.



Figure 18: Per-cycle routine monthly WBE laboratory workflow - Day 0 (field collection and receipt) through concentration, extraction, TAC screening, dPCR quantification, and metagenomic sequencing to Day 7 (report).

Quality Control Framework Summary

All laboratory processes are embedded with defined QC requirements, predefined acceptance criteria, and documented reprocessing paths on failure, as indicated in Fig. 19. We will input the MS2 process control (acceptance criterion: recovery $\geq 1\%$) into every batch as an NTC and a positive control (must be within $\pm 0.5 \log_{10}$ of the expected value). Batches that fail any checks are re-extracted, and two consecutive batch failures trigger a root-cause analysis and, where relevant, reagent-lot escalation to the supplier.

The decision-critical parameters of the laboratory pipeline are summarised here, with full step-by-step procedures and complete acceptance criteria in the referenced SOPs. TaqMan Array Card screening: a target is called positive at a cycle threshold (Ct) of ≤ 40 in at least one assay replicate, with Ct values of 38–40 re-tested and confirmed by dPCR; every card includes an extraction negative, a no-template control, and the MS2 positive control. Digital PCR: a reaction is accepted when it yields $\geq 8,000$ valid partitions and is called positive at ≥ 3 positive partitions; samples failing either criterion are re-run, and a second failure triggers re-extraction. Metagenomic sequencing: runs target ≥ 2 million quality-filtered reads per sample

(Nanopore flow cells are run until the target-coverage curve plateaus), and libraries must pass a $\geq 90\%$ read-quality filter for analysis; whole-genome sequencing of bacterial isolates targets $\geq 30\times$ mean coverage, and viral genomes $\geq 100\times$ at consensus positions. Across all batches, MS2 process-control recovery must be $\geq 1\%$ and the positive control within $\pm 0.5 \log_{10}$ of its expected value (Fig. 19; Supplementary Table S5).

a

Workflow stage	Control materials	Acceptance criterion	Action on failure
Sample receipt and storage	Temperature logger in the transport box; recorded sample volume	Box at 4–10 °C on arrival; sufficient volume for the assay (≥ 10 mL of swab eluate)	Reject and re-sample, or record the temperature deviation and re-test
Viral concentration	MS2 process control spiked into every sample before concentration	MS2 recovery $\geq 1\%$ of the spiked input; concentrated volume ≥ 0.5 mL	Re-concentrate the stored sample; if recovery stays low, optimise the elution step
Nucleic-acid extraction	Extraction negative control, no-template control and an external positive control	NTC and extraction negative undetected; positive control within $\pm 0.5 \log_{10}$ of the expected value; A260/280 ≥ 1.8	Re-extract the affected samples; investigate contamination; check the reagent lot
TaqMan Array Card screening	Extraction negative, no-template control and MS2 positive control on every card	Target called positive at Ct ≤ 40 in at least one replicate; Ct 38–40 re-tested and confirmed by dPCR	Re-run the card; confirm borderline calls by dPCR before reporting
Digital PCR quantification	No-template control and a positive control on every plate	Plate accepted at $\geq 8,000$ valid partitions; target called positive at ≥ 3 positive partitions	Re-run the plate; a second failure triggers re-extraction of the sample
Metagenomic sequencing	Library negative control and a mock community where available	≥ 2 million quality-filtered reads per sample; $\geq 90\%$ of reads pass the quality filter; coverage curve plateaus	Re-sequence the library; check loading concentration and flow-cell quality
Whole-genome sequencing	Negative library control and a reference strain where available	Bacterial genomes $\geq 30\times$ mean coverage; viral genomes $\geq 100\times$ coverage at consensus positions	Re-prepare and re-sequence; report the genome with its coverage caveat
Bioinformatics and reporting	Internal positive reference dataset and a version-pinned pipeline	Pipeline reproduces the reference output exactly; 100% concordance between replicate analyses; report format validated	Re-run the pipeline at the pinned version; hold the report until the discrepancy resolves

Batch-level rules

Every batch includes an extraction negative, a no-template control and an MS2 process control. A batch that fails any check is re-extracted; two consecutive batch failures trigger a documented root-cause analysis and, where relevant, reagent-lot escalation to the supplier.

Figure 19: Quality-control framework, mapped to each workflow stage: control materials, acceptance criteria, and defined corrective actions on failure

Ethics and Biosafety

Ethical considerations

We will ensure that the framework rests on the foundational premise that a pooled and anonymized community-level specimen is used within the Kenyan site. No individual samples will be collected, no individual identifiers will be generated, and no link will be created between a WBE signal and a specific

person. Thus, the design eliminates the primary individual-consent, privacy, and data-attribution concerns of clinical surveillance while retaining population-level intelligence. Additionally, we ensure that the following approvals and protections are in place:

- i. KEMRI Scientific and Ethics Review Unit (SERU) approval for the full protocol prior to field activation (reference to be inserted upon issuance).
- ii. The National Environment Management Authority (NEMA) environmental permit for the collection of wastewater samples at designated sites, together with guidelines for the safe disposal of analyzed sample materials.
- iii. National Commission for Science, Technology, and Innovation (NACOSTI) research licence for activities conducted in Kenya.
- iv. Written authorization from the County Director of Health in each participating county (Nairobi, Mombasa, Kisumu, Garissa, and Isiolo) before any site within their jurisdiction is activated.

All data will be managed under the Kenya Data Protection Act 2019 (data minimization, purpose limitation, storage limitation, and access authority) and overall will be guided by the principle of Helsinki. Only de-identified, aggregated data will be shared with international partners such as the WHO and Africa CDC. Individual-level clinical data from sentinel surveillance sites will also be de-identified before integration with WBE data. Finally, we will embed community engagement within the protocol: site-specific sensitization with community leaders and relevant authorities will precede site activation, and community-raised concerns will be documented and resolved before surveillance begins.

Governance and oversight. A standing National WBE Technical Working Group (TWG), convened by KNPPI with representatives of the implementing laboratories, county health management teams, and academic partners, will provide continuing oversight of the framework. The TWG will meet monthly during the first year and quarterly thereafter, and will be responsible for approving any change to analytic methods or alert thresholds before it takes effect; reviewing and versioning the statistical analysis plan; approving additions to, or retirement of, surveillance targets; reviewing network performance against its design assumptions, including the detection-sensitivity metrics described in Phase 5; and logging and reviewing community feedback forwarded by county teams. Substantive changes in scope (new pathogen targets, new site categories, or linkage with new clinical datasets) will require amendment of the KEMRI SERU approval and of county authorisations before implementation, and will be recorded in a change log summarised in the annual surveillance report. Requests for access to disaggregated or line-list data will be reviewed by a Data Access Committee against the Kenya Data Protection Act 2019, with all approvals and refusals documented; only de-identified, aggregated outputs will be released for public reporting. Community engagement will remain a standing agenda item for the TWG, and each annual report will summarise the concerns raised, the responses given, and any resulting protocol changes.

Gender and inclusion considerations

Because the collection matrix is pooled community wastewater, the framework inherently captures all genders, ages, and socioeconomic strata within the contributing population, a natural correction for the

gender and access biases documented in clinical surveillance. Field teams focus on gender balance, particularly where female field workers are better positioned to conduct community sensitization or access certain sampling locations. Training programs actively develop female scientists in participating laboratory teams, in line with KNPHI's institutional equity commitments. Clinical cross-reference data used to contextualize WBE signals are disaggregated by sex and age group, wherever available, to detect disproportionate disease burden in specific demographic groups.

Biosafety

All wastewater samples were handled as UN3373 Biological Substance Category B throughout the pre-analytical phase. Laboratory processing is BSL-2+: all open sample-processing steps occur inside a Class II Type A2 biosafety cabinet; PPE (double nitrile gloves, disposable gowns/coveralls, eye protection, and surgical/N95 respirators - N95 mandatory for all aerosol-generating steps including vortexing, sonication, and bead-beating) is worn at all times; and a unidirectional workflow (sample reception → extraction → PCR setup → PCR run → post-PCR analysis) is enforced with physically distinct zones and no reverse material movement. All liquid waste was inactivated with 10% sodium hypochlorite (minimum 30-minute contact) before drain disposal; all solid infectious waste (tips, tubes, extraction columns) was autoclaved at 121 °C for 30 min in biohazard autoclave bags before regulated-medical-waste disposal.

Processing for *M. tuberculosis* (Tier 3, exploratory) was restricted to the KNPHI BSL-3 certified laboratory. All personnel performing BSL-3 activities complete institutional BSL-3 training and undergo health surveillance (baseline tuberculin skin test or interferon-gamma release assay [IGRA]; annual re-testing) before commencing work. Any accidental exposure triggers immediate spill containment with 10% hypochlorite, personnel decontamination, occupational-health assessment within 4 h, and incident reporting within 24 h.

Sustainability Planning

For the WBE system to deliver lasting value beyond the initial grant period, a concurrent sustainability strategy is implemented alongside scientific activities. Its four pillars - institutional ownership, financial sustainability, workforce continuity, and technology stewardship - are summarised in Fig. 20.

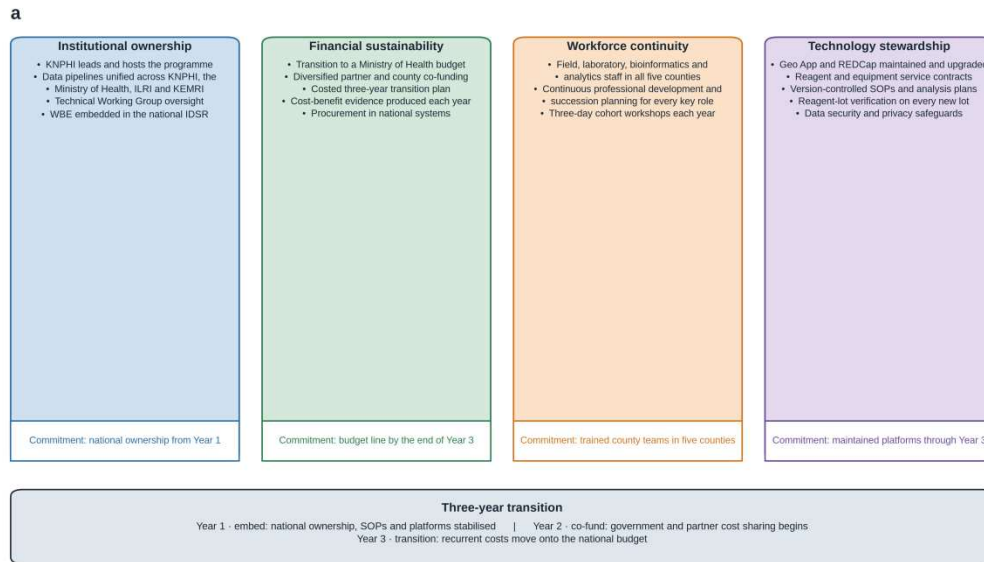


Figure 20:Sustainability planning framework - the four pillars and the specific commitments Kenya has undertaken under each

Budget

The total study cost for the three-year framework was USD 1,021,400 in direct and institutional funds, allocated strategically across technical operations, capacity building, and transport logistics (Fig. 21).

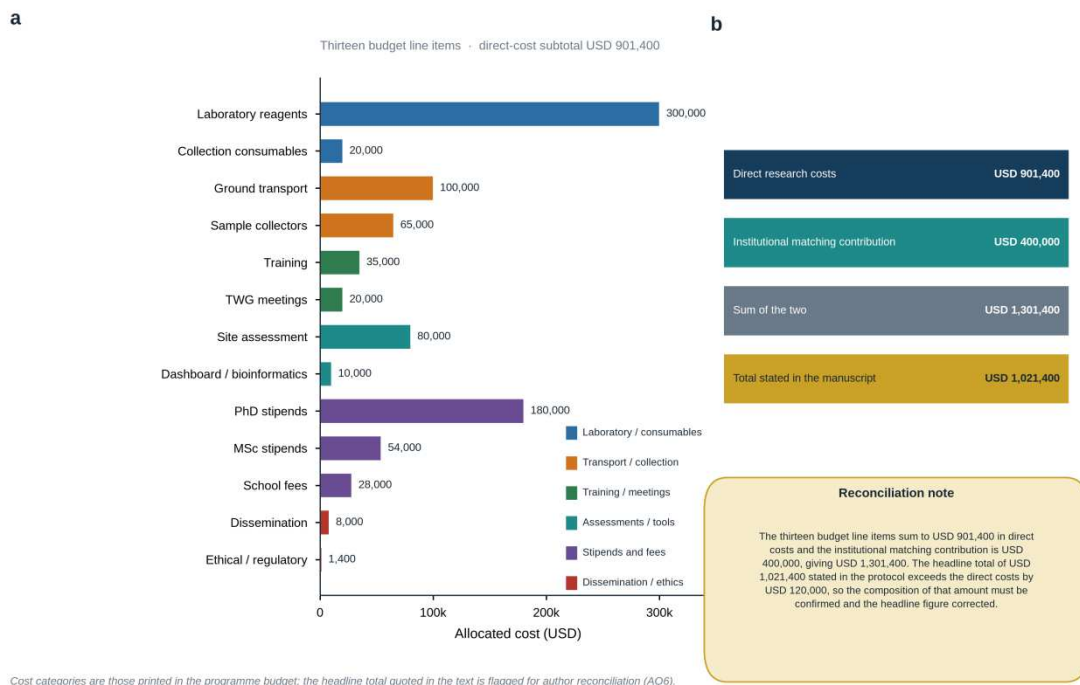


Figure 21:Budgetary allocation for the three-year implementation of the Kenya national WBE protocol (total USD 1,021,400).

References

1. Kabudula, C. W. *et al.* Progression of the epidemiological transition in a rural South African setting: findings from population surveillance in Agincourt, 1993–2013. *BMC Public Health* **17**, (2017).
2. Gahimbare, L. *et al.* Antimicrobial Resistance in the WHO African Region: A Systematic Literature Review 2016–2020. *Antibiotics* vol. 13 Preprint at <https://doi.org/10.3390/antibiotics13070659> (2024).
3. Osen, G. *et al.* Antimicrobial resistance in Africa: A retrospective analysis of data from 14 countries, 2016–2019. *PLoS Med.* **22**, (2025).
4. Koua, E. L. *et al.* Trends in public health emergencies in the WHO African Region: An analysis of the past two decades public health events from 2001 to 2022. *BMJ Glob. Health* **8**, (2023).
5. Fall, I. S. *et al.* Integrated Disease Surveillance and Response (IDSR) strategy: Current status, challenges and perspectives for the future in Africa. *BMJ Global Health* vol. 4 Preprint at <https://doi.org/10.1136/bmjgh-2019-001427> (2019).
6. Khan, M. A. *et al.* Mathematical analysis of typhoid model with saturated incidence rate. *Adv. Stud. Biol.* **7**, (2015).
7. Ali, M., Nelson, A. R., Lopez, A. L. & Sack, D. A. Updated global burden of cholera in endemic countries. *PLoS Negl. Trop. Dis.* **9**, (2015).
8. Sims, N. & Kasprzyk-Hordern, B. Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. *Environment International* vol. 139 Preprint at <https://doi.org/10.1016/j.envint.2020.105689> (2020).
9. Daughton, C. G. Illicit drugs in municipal sewage: Proposed new nonintrusive tool to heighten public awareness of societal use of illicit-abused drugs and their potential for ecological consequences. *ACS Symposium Series* **791**, (2001).
10. Kirby, A. E. *et al.* Using Wastewater Surveillance Data to Support the COVID-19 Response — United States, 2020–2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, (2021).
11. Antimicrobial consumption and resistance in bacteria from humans and food-producing animals: Fourth joint inter-agency report on integrated analysis of antimicrobial agent consumption and occurrence of antimicrobial resistance in bacteria from humans and food-producing animals in the EU/EEA JIACRA IV – 2019–2021. *EFSA Journal* **22**, (2024).
12. Sosa-Hernández, J. E. *et al.* Extensive Wastewater-Based Epidemiology as a Resourceful Tool for SARS-CoV-2 Surveillance in a Low-to-Middle-Income Country through a Successful Collaborative Quest: WBE, Mobility, and Clinical Tests. *Water (Switzerland)* **14**, (2022).
13. Strike, W. *et al.* Implementing Wastewater-Based Epidemiology for Long-Read Metagenomic Sequencing of Antimicrobial Resistance in Kampala, Uganda. *Microorganisms* **13**, (2025).
14. Mtetwa, H. N., Amoah, I. D., Kumari, S., Bux, F. & Reddy, P. Surveillance of multidrug-resistant tuberculosis in sub-Saharan Africa through wastewater-based epidemiology. *Heliyon* **9**, (2023).
15. Mtetwa, H. N., Amoah, I. D., Kumari, S., Bux, F. & Reddy, P. Exploring the role of wastewater-based epidemiology in understanding tuberculosis burdens in Africa. *Environ. Res.* **231**, (2023).
16. Dzinamarira, T. *et al.* Epidemiological surveillance of enteric viral diseases using wastewater in Africa – A rapid review. *Journal of Infection and Public Health* vol. 15 Preprint at <https://doi.org/10.1016/j.jiph.2022.05.010> (2022).
17. Tiwari, A. *et al.* Strengthening pathogen and antimicrobial resistance surveillance through environmental monitoring in sub-Saharan Africa: stakeholder perspectives. *Int. J. Hyg. Environ. Health* **270**, (2025).
18. Nkengasong, J. N. & Tessema, S. K. Africa Needs a New Public Health Order to Tackle Infectious Disease Threats. *Cell* vol. 183 Preprint at <https://doi.org/10.1016/j.cell.2020.09.041> (2020).

19. Principi, N., Autore, G., Ramundo, G. & Esposito, S. Epidemiology of Respiratory Infections during the COVID-19 Pandemic. *Viruses* vol. 15 Preprint at <https://doi.org/10.3390/v15051160> (2023).
20. Lee, H., Kim, J., Kim, J. & Park, Y.-J. Review of the Global Burden of Tuberculosis in 2023: Insights from the WHO Global Tuberculosis Report 2024. *Public Health Weekly Report* **18**, (2025).
21. Galkissa-Dewage, B. *et al.* Monitoring SARS-CoV-2 in wastewater in Kisumu, Kenya: Association with COVID-19 case numbers. *PLOS Water* **4**, (2025).
22. Sichamba, P. *et al.* A Retrospective Analysis of Wastewater Confirms Dominant Circulation of SARS-CoV-2 Delta Variant in Nairobi, Kenya, between April 2021 and August 2021. *Am. J. Mol. Biol.* **12**, (2022).
23. Hendriksen, R. S. *et al.* Pathogen surveillance in the informal settlement, Kibera, Kenya, using a metagenomics approach. *PLoS One* **14**, (2019).
24. Shafagati, N. *et al.* Enhanced detection of respiratory pathogens with nanotrap particles. *Virulence* vol. 7 Preprint at <https://doi.org/10.1080/21505594.2016.1185585> (2016).
25. Keenum, I. *et al.* Optimizing Wastewater Surveillance: The Necessity of Standardized Reporting and Proficiency for Public Health. *Am. J. Public Health* **114**, (2024).
26. Servetas, S. L. *et al.* Standards to support an enduring capability in wastewater surveillance for public health: Where are we? *Case Studies in Chemical and Environmental Engineering* **6**, (2022).

Supplementary Information

The following supplementary items accompany this protocol and are delivered as separate files:

- i. Supplementary File S1 - Pathogen Priority Scoring Matrix Calculator (.xlsx). Full 64-pathogen calculator with country reference data (54 countries), seven-criterion weights, endemicity multipliers, threshold table, feasibility gate, and per-country data source provenance.
- ii. Supplementary Table S1 – Ranked composite scores for all 64 candidate pathogens in Kenya (pilot pathway).
- iii. Supplementary Table S2 - Activity-level Gantt chart, 2026-2029.
- iv. Supplementary Table S3 Full TAC assay layout, primers, probes, and cycling parameters.
- v. Supplementary Table S4 - Site-level metadata for all 57 activated sites.
- vi. Supplementary Table S5 - Quality-control acceptance criteria per workflow stage.
- vii. Supplementary Table S6 - Reagent lot-verification checklist.
- viii. Supplementary Documents S1–S12 – Standard Operating Procedures.
- ix. Supplementary Form F1 - Standardised sample collection metadata form.
- x. Annex A - Kenya WBE Minimum Data Set (MDS) full data dictionary and eIDSR / DHIS2 API mapping.
- xi. Annex B - Community sensitisation and consent-of-locality template.
- xii. Annex C - Alert-tier escalation SOP and SitRep template (English and Kiswahili).
- xiii. Supplementary Document S13 - Pre-specified statistical analysis plan (sample-size and detection-power calculations, AI/ML validation procedures, and trend-state classification rules).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTablesS1S7.xlsx](#)
- [pathogenPrioritisation.xlsx](#)