

Supporting Information

People's Water Data: Design and Validation of a Scalable Decentralized Framework for Monitoring Drinking Water Safety

Suzan Kagan^{1,2}, Sankar Sudhir¹, Ben Hamilton³, Alona Maslennikov^{2,4}, P. Amaravathy⁵, Timor Lichtman⁶, Sonali Seth¹, Tanmayaa Nayak¹, Ramya Dwivedi¹, Ganesan Velmurugan^{5,7}, Hadas Mamane², Ram Fishman⁶ and Thalappil Pradeep^{*1,8}

¹DST Unit of Nanoscience and Thematic Unit of Excellence, Department of Chemistry, IIT Madras, Chennai 600 036, India.

²School of Mechanical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel.

³Faculty of Environment, Science and Economy, University of Exeter, Exeter EX4 4QF, United Kingdom.

⁴School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

⁵Chemomicrobiomics Laboratory, Department of Biochemistry and Microbiology, KMCH Research Foundation, Coimbatore 641 014, Tamil Nadu, India.

⁶Department of Public Policy and New Environment School, Tel Aviv University, Tel Aviv 6997801, Israel.

⁷Central Research Laboratory, KMCH Institute of Health Science and Research, Coimbatore 641 014, Tamil Nadu, India.

⁸International Centre for Clean Water, 2nd Floor, B-Block, IIT Madras Research Park, Kanagam Road, Taramani, Chennai 600113, India

*Email: pradeep@iitm.ac.in

Tel.: +91-44 2257 4208; Fax: +91-44 2257 0545/0509

35 **Table of Contents**

36 1. Theoretical and operational training framework.....3

37 2. Household survey tool development.....6

38 .3 Sampling strategy, geospatial design, and allocation refinement15

39 4. Ethical committee approval29

40 5. Laboratory analysis and measurements29

41 6. Health survey tool35

42 7. Household feedback.....37

43 8. Implementation scale and household-level monitoring density.....40

44 9. Variability in survey duration40

45 10. Agreement rates41

46 11. Data quality risks and mitigation45

47 12. Resource-efficiency and monitoring-density implications48

48

49

1. Theoretical and operational training framework

The theoretical and operational foundation supporting the People's Water Data (PWD) monitoring framework was established through the National Program on Technology Enhanced Learning (NPTEL), a Government of India-supported digital learning initiative offering open-access courses developed by the Indian Institutes of Technology (IITs) and the Indian Institute of Science (IISc), India's leading public institutions for engineering and scientific research. The course titled "A Hybrid Course on Water Quality - An Approach to People's Water Data" (IIT Madras Course ID 5011) was offered by the Department of Chemistry at IIT Madras, in collaboration with Tel Aviv University (Israel) and KMCH Research Foundation (India), and functioned as a protocol-alignment and standardization mechanism for large-scale household-level surveillance activities and as a structured capacity-building mechanism for decentralized monitoring implementation.

The course integrated structured online lectures, laboratory demonstrations, and supervised field assignments to ensure methodological consistency and analytical reliability across geographically distributed monitoring locations. The curriculum was organized into six core modules integrating water science, analytical measurement, behavioral assessment components, and sustainability frameworks relevant to drinking water resource management. Evaluation combined theoretical assessment with structured field deployment over four weeks of supervised monitoring activities. Participants were trained in household interaction protocols, the measurement of physicochemical and microbiological parameters, digital data capture, and standardized integration of field data into the PWD digital data architecture.

Learning objectives emphasized operational accuracy in water quality assessment, structured survey execution, digital data handling, and analytical interpretation within sustainable water management context. Complementary modules addressed survey design, water microbiome considerations, measurement reliability, and behavior-linked risk assessment to strengthen interdisciplinary integration within the monitoring framework.

74 The course content was jointly developed and delivered by experts in water science, environmental
 75 engineering, and sustainability. The core faculty included Prof. Thalappil Pradeep and Prof. Ligy Philip
 76 from the Indian Institute of Technology Madras, Prof. Hadas Mamane (Tel Aviv University and Adjunct
 77 Faculty at IITM), Prof. Ram Fishman (Tel Aviv University), and Dr. G. Velmurugan (KMCH Research
 78 Foundation). To complement the core instruction, a series of guest lectures introduced specialized technical
 79 domains and applied perspectives relevant to the implementation of decentralized surveillance. The
 80 sessions covered topics including water governance and water management, geospatial data science, water
 81 entrepreneurship and innovation, radiological parameters in water quality, agrochemical contamination,
 82 water microbiome and microbial ecology, behavioral obstacles and the role of enforcement agencies in
 83 water quality management. These components collectively supported a multidisciplinary operational
 84 understanding of water quality monitoring, integrating scientific, analytical, behavioral, and regulatory
 85 dimensions relevant to structured drinking water surveillance systems. A detailed syllabus including
 86 lecture structure, learning objectives, reference literature, and faculty contributors is presented in **Table**
 87 **S1.**

88 **Table S1.** Syllabus of the academic course (IIT Madras Course ID 5011) titled “*A Hybrid Course on Water Quality - An*
 89 *Approach to People’s Water Data.*”

Category	Details
Course Title	Water Quality Monitoring and People-Centered Water Data Systems
Course Description	The course provides theoretical and operational training in drinking water quality assessment, decentralized monitoring systems, and community-linked data generation. It integrates foundational water science, analytical measurement techniques, behavioural dimensions, hydroinformatics, and supervised field deployment to support standardized large-scale monitoring.
Course Delivery	Structured lectures, laboratory demonstrations, guided discussions, supervised field measurements, and data analysis exercises.
Evaluation Method	Theoretical assessment combined with four weeks of supervised household-level water quality monitoring and structured data integration.

Learning Outcomes	Identify major drinking water contaminants and associated health risks; interpret national and international drinking water standards; measure physicochemical, microbiological, and radiological parameters; design and implement field-based surveys; integrate geospatial data; apply hydroinformatics principles; digitize and manage datasets; perform statistical analysis and modelling; assess behavioural and risk-related dimensions of household water use.
Module A: Foundations of Water Quality	Water cycle; potable water concepts; microbial, chemical, heavy metal, radiological, and particulate contamination; measurement units; accuracy and precision; sampling protocols.
Module B: Water Quality Parameters and Standards	Core parameters (temperature, pH, conductivity, dissolved oxygen, alkalinity, oxidation–reduction potential, turbidity); microbial contamination; chemical contaminants (ammonia, nitrate, chloride, BOD, TOC); agrochemicals; emerging contaminants (pharmaceuticals, personal care products, PFAS); heavy metals (including arsenic); radiological parameters; regulatory frameworks and enforcement; microbial diversity and antimicrobial resistance.
Module C: Measurement Techniques and Instrumentation	Instrumentation for water quality analysis; calibration and validation; quality control procedures; heavy metal analysis; ultra-trace measurements; disinfection measurement methods; wastewater contamination risk assessment.
Module D: Field Survey and Hydroinformatics	Field test kits (FTKs); geospatial data acquisition and integration; survey research design; statistical analysis of water data; household and rural purification technologies; introduction to hydroinformatics and digital data systems.
Module E: Measurement, Behaviour, and Public Health Impact	Household water practices; behavioural barriers to improving water quality; role of measurement and information in influencing behaviour.
Module F: Supervised Field Deployment	Field training in measurement techniques; structured data collection; quality assurance; statistical analysis and interpretation; innovation using community-generated water quality data.

2. Household survey tool development

The survey component of the People's Water Data (PWD) monitoring framework was iteratively refined to enhance reliability, technological compatibility, and scalability (Table S2).

Table S2. Phased development and validation of the household water survey platform

Phase	Platform	Location / Period	Coverage / Sample Size	Key Features	Data Processing and Validation
Pilot Phase	Kobo Toolbox	Nallampatti village, Tamil Nadu	262 households; 150 rural water sites	Initial structured digital survey deployment	Pilot testing of survey structure and field feasibility
Expanded Deployment	Qualtrics	January–May 2024	1,735 total surveys (household and source water)	Branching logic; time tracking for reliability testing; automatic geolocation tagging	Refined to 1,054 validated entries after data cleaning and duplicate removal; 610 geotagged records retained; cross-verification between household responses, field images, and laboratory results; real-time multi-institutional data synchronization
Scaled Implementation	EpiCollect5 (open-source mobile platform)	Subsequent national and cross-regional deployments	Scalable multi-institutional implementation	Online and offline data collection; automatic GPS stamping; direct	Standardized survey structure retained across versions emphasizing traceability and

				integration with PWD dashboard; enhanced Android device compatibility	participant consent; fully integrated with the PWD data architecture to support automated synchronization and streamlined field workflows
--	--	--	--	---	---

The household water survey was developed as a comprehensive instrument to capture both the objective and subjective dimensions of drinking water quality at the point of use. As detailed in **Table S3**, the survey systematically collected information on water sources, household demographics, and practices related to storage, treatment, and usage. It distinguished between natural sources (such as rivers, lakes, or waterfalls) and artificial sources (such as wells, bore wells, and municipal taps), ensuring representation across diverse supply systems.

Additional sections addressed household composition, education levels, and water-handling behaviors, followed by questions on storage practices, treatment methods, and frequency of cleaning or maintenance. Modules on health impacts explored self-reported waterborne illnesses and their economic implications, while sections on perceived quality and supply reliability captured user confidence and satisfaction with their drinking water.

Standardized questions balanced clarity with scientific rigor, combining closed-ended responses and observational checks to minimize reporting bias. Field teams complemented survey responses with on-site testing including 5-in-1 dip strips for chemical parameters (hardness, alkalinity, chlorine, pH) and *E. coli*/coliform kits for microbial detection alongside photographic documentation and GPS tracking to ensure data traceability. This integrated design linked physicochemical and microbiological measurements with socio-behavioral context, strengthening the interpretability and applied relevance of the dataset.

113 **Table S3.** Comprehensive survey tool for assessing household water quality, practices, and perceptions.

Question	Options	Question type
Location (Click 'Update Location' and ensure the accuracy)		GPS
Informed Consent: I agree to participate in this study	Yes; No	MCQ
Sample ID (Example: XX01091225 (initials-day-month-exact time))		Typing it
Time (Autoset; click next)		Autoset
Type of Water Sample		MCQ
Photo of Street Overview: Take a photo of the street front		Photo
Photo of House and Surrounding: Photograph the target area		Photo
Photo of House Gate and Number: Ensure a clear photo		Photo
Name of respondent		Typing it
Mobile Number		Typing it
Respondent Email ID		Typing it
Gender	Male; Female	MCQ
Age Group	Under 18; 18-24 years; 25-34 years; 35-44 years; 45-54 years; 55-64 years; 65+ years	MCQ
Position in your family	Head of household; Spouse of Head of household; Daughter/Son; Other family member	MCQ
Highest educational qualification	Literate but no Formal Schooling; SSC/HSC; Some	MCQ

	College but not Graduate; Graduate/Postgraduate (General); Graduate/Postgraduate (Professional)	
Number of children under the age of 5 in your home		Typing it
Time (Autoset; click next) — after Household info		Autoset
Time (Autoset; click next) — start of Section II.A		Autoset
Where do they fetch the glass of drinking water from?	Tap (pipeline); Any kind of container	MCQ
Take a photo of the container/ pot/ RO Unit from which water was fetched		Photo
Specify the type of container from which they fetch water	Plastic container; Clay pot; Metal container	MCQ
Did the respondent fingers touch the water while they are filling the cup?	Yes; No	MCQ
How they fetched the water from the container	Pouring directly from the container; Using container's tap; Using a dedicated cup or a jug; Other	MCQ
Time (Autoset; Just touch next) — end of Section II.A		Autoset
Time (Autoset; click next) — start of Section II.B		Autoset
Do you wash the dedicated cup?	Yes; No	MCQ
How often do you wash the dedicated cup?	After every use; Once a day; Don't know; Other	MCQ

Was the drinking water ever stored in a container?	Yes; No	MCQ
What material is the container is made of?	Plastic; Clay; Aluminum; Other	MCQ
What is the size of the container?	1-5L; 6-10L; 11-20L; 21-50L; 51-70L; 71-100L; >100	MCQ
Is the container covered?	Yes; No	MCQ
Is the container placed less than 2 meters away from cooking environment?	Yes; No	MCQ
Is the container on the ground?	Yes; No	MCQ
When did you last clean the container?	Every day; This week; This month; Last 2 months or more; This year; Never; Don't know	MCQ
Did you use any detergents for cleaning your container?	Yes; No	MCQ
Time (Autoset; click next) — start of Source section		Autoset
Do you use different types of water for drinking and bathing/washing dishes?	Yes; No	MCQ
How does your household get its drinking water supply?	Fetch the water by myself; By pipeline; I Purchase the drinking water	MCQ
From where did you fetch the drinking water?	Communal tap; Self collection from River/Lake/Stream; Borewell water / Ground water; Well	MCQ
What is the source of piped water?	Municipal piped water supply (Metro water); Borewell water / Ground water; Well	MCQ

What kind of drinking water did you purchase?	Canned water (not branded); Water filled from a truck / lorry; Packaged water (branded)	MCQ
Specify the type of well	Well within my property; Community well	MCQ
Location of well		Typing it
Is the well covered?	Yes; No	MCQ
Has your drinking water stored in a roof tank?	Yes; No	MCQ
When was the last time you cleaned your roof tank?	Within the last 3 months; Within the last 6 months; Within the last year; More than a year ago or never; Can't remember	MCQ
Time (Autoset; just click next) — start of Treatment		Autoset
Do you treat your water in anyway?	Yes; No	MCQ
How did you treat today's drinking water?	Filter unit / RO; Boiled; Strained using a cloth; Added material	MSQ
How long boiling of water is performed?	Less than 1 minute; 1 minute; 2 minutes; More than 2 minutes; Until the water starts bubbling; Don't know	MCQ
When did you last boil water?	Today; Yesterday; 2 days ago or more	MCQ
Do you wash the cloth that you use for straining	Yes; No	MCQ
How often do you replace/ wash the cloth?	After every use; Once a day; Once every couple of days;	MCQ

	Once a week; Don't know; Other	
What kind of filter unit do you have?	RO (Reverse Osmosis); Ceramic / Clay; UV; Screen filter; Activated Carbon; Sand; Composite; Don't know; Other	MCQ
When was the last time the RO/ filter unit maintenance service arrived?	In the last 3 months; In the last 6 months; In the last year; My RO system is less than 3 months old; Never	MCQ
Mention the brand and model of the RO/ Filter Unit		Typing it
Please take a picture of your Filter unit/ RO Unit		Photo
Is anything done for maintaining the RO/ filter unit?	Yes; No	MCQ
What does the maintenance includes?	Replacing parts; Running a backwash; Sending to a private company; A private company comes to service the machine; Don't know; Other	MSQ
What is the maintenance time for your RO unit mentioned in the manual?	Every day; Every week; Every 6 months; Once a year; Don't know; Other	MCQ
Do you maintain your filter unit/ RO on the required time?	Yes; No	MCQ
Time (Autoset; click next) — end of Treatment		Autoset
Time (Autoset; Just touch next) — start of Assessment		Autoset

Please take a cup of your drinking water, what is the color?		MCQ
What does your drinking water taste like?		MCQ
What does your drinking water smell like?		MCQ
How do you feel about the safety of this water before anything is done to it?	Very bad; Bad; Satisfactory; Good; Very good	MCQ
What is the risk of becoming ill by consuming your water?	Very risky; Some risk; Little risk; No risk	MCQ
Have you or any of your family members suffered during last month?	Diarrhea; Sore throat; Muscle pain; Change in teeth health; Malaria; Joint pain; Bone health issues; None of the above; Other	MSQ
What is the last time this household didn't have enough drinking water?	This month; This year; Before this year; Never	MCQ
Select the type of water body	Open well; Bore well; Pond; Pumping station; Piping system; Public tap; River; Lake; Other	MCQ
Local name of the water body (if private well mention the identification)		Typing it
Utilization purposes of this water source (check all which are relevant)	Drinking (Humans); Drinking (cattles); Bathing; Irrigation; Fishing; Not being used; Not aware; Other purposes	MSQ
Does people bath/ swim in this water body?		MCQ

Did you notice discharge of sewage or waste water into the water body?	Yes; No	MCQ
Did you notice plastics (either floating or in the sediments)?		MCQ
Does the water body has more algal blooms? (greenish water)		MCQ
Does the water body has any floating aquatic plants?		MCQ
Does the water body has fishes and other aquatic animals?		MCQ
Did you notice any religious activity like disposal of flowers etc.?		MCQ
Presence of farming lands within 1 km radius from collection point		MCQ
Presence of industries within 1 km radius from collection point		MCQ
Take a picture of the water source		Photo
Time (Autoset; click next) — end of Assessment		Autoset
Date (Autoset; click next) — Water Parameters		Autoset
Time (Autoset; click next) — Water Parameters		Autoset
Sample No. — Water Parameters		Typing it
Total Chlorine (mg/L)	0; 0.5; 1; 2; 4; 10	MCQ
Free Chlorine (mg/L)	0; 0.5; 1; 2; 4; 10	MCQ
Hardness (mg/L)	0; 25; 50; 120; 250; 425	MCQ

Alkalinity (ppm)	0; 40; 80; 120; 180; 240	MCQ
pH — sample reading	6.2; 6.8; 7.2; 7.8; 8.4	MCQ
Turbidity value 1 (NTU)		Typing it
Turbidity value 2 (NTU)		Typing it
Turbidity AVG (NTU)		Typing it
pH (duplicate field)	6.2; 6.8; 7.2; 7.8; 8.4	MCQ
ORP (mV)		Typing it
Select the unit for TDS		MCQ
TDS (ppt)		Typing it
TDS (ppm)		Typing it
Select the unit for Conductivity		MCQ
Conductivity (μ S)		Typing it
Conductivity (mS)		Typing it
Date (Autoset; click next)		Autoset
Biological tests		
Time (Autoset; click next)		Autoset
Biological tests		
Sample No. Biological tests		Typing it
Presence of <i>E. coli</i>	Positive; Negative	MCQ
Presence of Total Coliforms	Positive; Negative	MCQ
Photo of Bio Kit		Photo

3. Sampling strategy, geospatial design, and allocation refinement

Household sampling was implemented using a polygon-based spatial allocation approach rather than unrestricted convenience sampling. For each study area, polygons were defined by the research team, course instructors, or local project coordinators to represent the geographic area of interest for household-level water-quality assessment. Polygon boundaries were established using GPS coordinates, local operational knowledge, and field-feasibility considerations. These polygons constrained sampling to predefined spatial boundaries and provided a consistent operational unit for assigning household surveys across study areas.

123

124 Within each polygon, participants were instructed to complete ten valid household surveys at distinct
125 locations. Therefore, the target number of household assessments within a polygon or deployment area
126 was determined by the number of participating participants multiplied by ten. This structure standardized
127 participant workload while allowing the overall number of household assessments to scale according to
128 the number of trained field operators assigned to each polygon or region.

129

130 The sampling protocol evolved through successive implementation phases. In early deployments,
131 participants selected households within assigned polygons while following geographic guidance intended
132 to reduce clustering and convenience-based selection. Training materials emphasized the risk of biased
133 household selection and the importance of distributing sampling effort across the assigned area rather than
134 concentrating surveys near highly accessible locations, such as campuses, main roads, familiar
135 neighborhoods, or the participant's starting point. Post-field GPS review from these early deployments
136 indicated that participant-selected locations could still cluster around accessible or familiar areas, despite
137 the use of general spatial guidance.

138

139 In later deployments, this process was strengthened by providing GIS-supported candidate coordinates
140 within the assigned polygon. Candidate sampling points were generated using a Sobol sequence-based
141 procedure to produce low-discrepancy, spatially distributed locations within the geographic area of interest.
142 Where needed, a bounding box was first defined around the study polygon, candidate points were generated
143 within that bounding box, and points falling outside the polygon were excluded until a sufficient number
144 of candidate locations was obtained for the required number of household surveys. The resulting
145 coordinates were used to guide field deployment rather than to define fixed household addresses. This
146 approach helped standardize sampling guidance across study areas, reduce participant-driven selection
147 bias, and improve reproducibility across repeated deployments.

148

149 An additional advantage of the Sobol sequence approach is that future deployments can generate additional
150 candidate points within the same polygon without intentionally targeting previously sampled locations,
151 thereby allowing household-level spatial resolution to increase over time. The goal was not to eliminate
152 field flexibility, but to guide participants toward better spatial dispersion and reduce the likelihood that
153 surveys would be concentrated near easily accessible streets, familiar areas, or the participant's starting
154 point.

155

156 Three sampling or allocation modes were used or considered during protocol development. The first mode
157 was participant-selected sampling within assigned polygons, used during early implementation phases to
158 test feasibility under real field conditions. The second mode was non-clustered allocation of Sobol-
159 generated candidate points to participants. In this approach, candidate points were spatially distributed
160 across the polygon, but individual participants could be assigned locations that were widely separated from
161 one another. This maintained polygon-level dispersion but could increase expected travel burden and route
162 crossing during field implementation. The third mode was K-means-supported allocation, in which nearby
163 Sobol-generated candidate points were grouped into participant-level spatial clusters. This approach was
164 adopted to make the ten-site field assignment more operationally feasible while maintaining broader
165 polygon-level coverage.

166

167 In both non-clustered and K-means-supported allocation, the candidate sampling points were generated
168 using the same Sobol-based method; what differed was how those points were assigned to individual
169 participants or field teams. Sobol-based point generation prioritized spatial dispersion and reproducibility,
170 whereas K-means-supported allocation prioritized logistical feasibility by grouping nearby candidate
171 locations. Because K-means clustering defines spatial clusters but does not inherently guarantee exactly

ten points per participant, clustered coordinates were treated as field guidance, and each participant was still required to complete ten valid household surveys.

When a suggested household was unavailable, declined participation, or could not be accessed safely, participants were instructed to identify the nearest eligible neighboring household within the same polygon. This substitution rule preserved the polygon constraint while allowing necessary field flexibility under real-world conditions. In such cases, the substituted household replaced the suggested coordinate for operational purposes, but the survey remained subject to the same documentation and quality-control requirements. All substitutions were recorded using the digital survey workflow, including GPS coordinates, timestamps, photographic documentation where appropriate, and the unique sample ID.

To reduce selection bias, participants were assigned to defined polygons or, where pre-generated allocation was used, to distinct candidate locations or spatial subareas. This distributed collective sampling effort across the polygon and reduced the likelihood of repeated sampling around highly accessible locations. Submitted records were reviewed using geolocation, timestamps, photographs, sample identifiers, and survey metadata to identify spatial clustering, incomplete coverage, unusual movement patterns, or potential protocol deviations.

Across countries and phases, the underlying principle remained consistent: household sampling was geographically constrained within predefined polygons and supported by spatial guidance to reduce convenience-based selection. However, the level of pre-generated coordinate guidance and allocation structure differed by deployment phase and local feasibility. Early deployments relied more heavily on flexible participant-selected sampling within assigned polygons, whereas later deployments incorporated stronger GIS-supported allocation, Google Maps guidance, Sobol-generated candidate locations, and K-means-supported grouping where feasible. Thus, protocol refinement moved from flexible participant-

selected sampling toward stronger GIS-guided allocation, while retaining field-level substitution when households were unavailable, inaccessible, or unsafe to approach.

Overall, the refined protocol was designed to improve spatial dispersion within assigned polygons, reduce repeated sampling near highly accessible locations, balance spatial coverage with field practicality, and strengthen the reproducibility of household-level sampling across study regions.

To ensure methodological consistency, data integrity, and operational safety during field deployment, all field operators adhered to a standardized protocol titled “Standard Operating Procedure (SOP) for Water spatially Guided Geographical Diversity, Sampling and Verification” as detailed in Text S3. This document, approved and periodically updated by the project coordinators, provides structured, step-by-step procedures for spatially guided and geographically distributed household and source-water sampling within defined spatial boundaries.

Objective: To establish a practical and reliable protocol for collecting household water samples and verifying results. This ensures high-quality, trustworthy data for understanding water quality and guiding participant’s interventions.

Scope: This SOP applies to all participants engaged in water sample collection, field measurements, and data verification during activities conducted under the PWD framework.

Spatially guided geographical diversity: Each participant (or field team) shall collect 10 water samples, which may include: (1) Household drinking water samples, and/or (2) Natural or source water samples (e.g., rivers, lakes, wells, borewells, ponds). The locations must be coordinated according to the instructions Use a spatially guided sampling method to select households:

Define the study area: Define the study area by dividing the city into predefined geographic regions (e.g., North Chennai, Central Chennai, South Chennai, West Chennai, and Suburban Areas) and enumerating all neighborhoods within each region using numerical identifiers. For example, in South and East Chennai, neighborhoods may be listed as follows: 1. Teynampet; 2. Thousand Lights; 3. Gopalapuram; 4. Mylapore;

5. Chromepet; 6. Egmore; 7. Chetpet; 8. Perungudi; 9. Sholingnallur; 10. Adyar; 11. Besant Nagar; 12. T. Nagar; 13. Thiruvanmiyur; 14. Saidapet; 15. Velachery; 16. Taramani; and additional neighborhoods as applicable. The complete list should be prepared in advance using Wikipedia, with a printed and marked map. Spatially guided select neighborhoods. For example, Group 1 will go to Neighborhood 15, this can be done manually printed as example (Figure S1 for an example of spatially guided allocation):

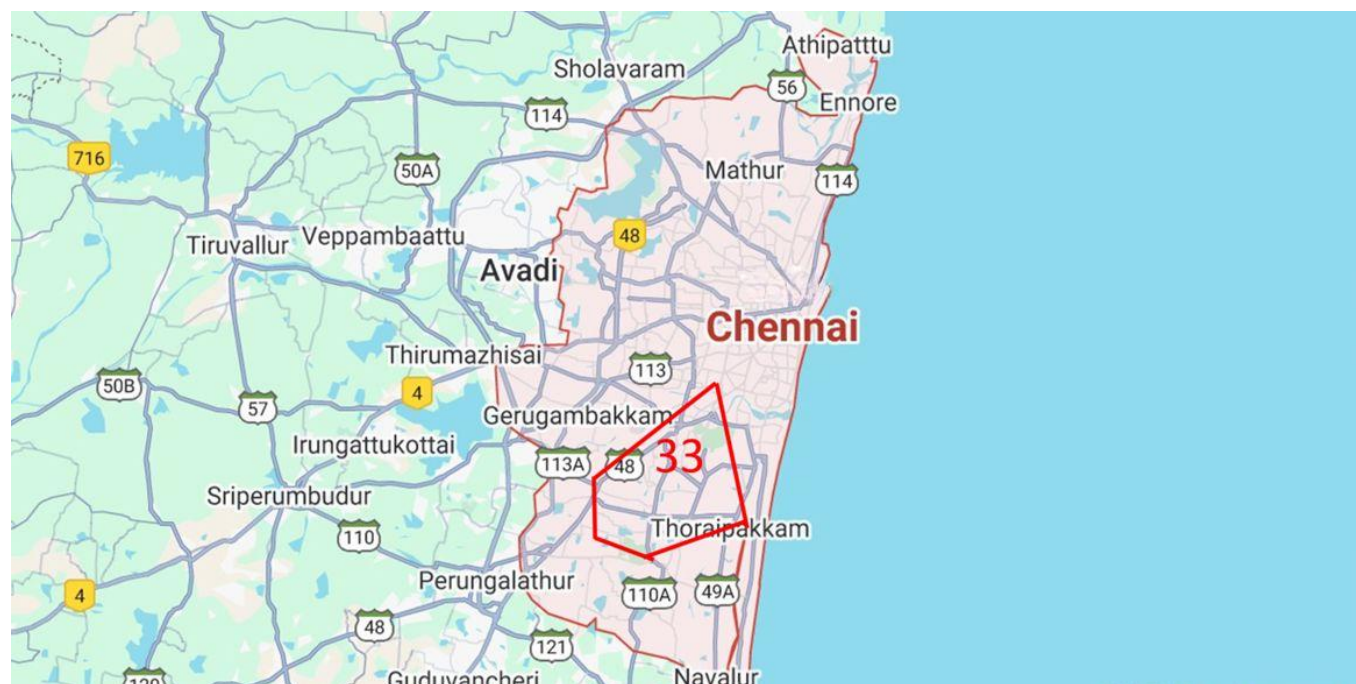


Figure S1: Spatially Guided Selected Sampling Zone (Zone 33) within Chennai Metropolitan Region

Field Navigation and Sampling Implementation Guidelines:

- If you are traveling from the lab or home toward the neighborhood, start by choosing one point from google map you received. Head toward this point first; it will serve as your base to orient yourself within the neighborhood and begin the selection process.
- For household sampling, enforce a minimum distance of 10 buildings between each sampling location to ensure spatial diversity.

- Confirm Participation: Once households are selected, check if they are accessible and willing to participate in the study.
- Use a systematic sampling method starting at a spatially guided household and then selecting every *n*th household (such as every 3rd or 5th) until you reach 10.
- For open water sources, implement a spatially guided selection algorithm that considers at least a 1 km distance between locations.
- No shops and no commercial water.
- Download a geo-tagged application named [app name: GPS Map Camera] and save photos on your phone for future reference. For backup.

Or **automatically**, a Google Maps link with potential locations will be shared with the participants. Once a participant reaches a location, they will mark it as completed, and it will be removed from the map to prevent other participants from same batch and day visiting the same household.

For example: https://www.google.com/maps/placelists/list/7iMEQOtiQ5Gx_Mvo3gPicw?g_st=iw

Phase-specific implementation and allocation refinement:

The same polygon-based sampling logic was used across study regions, but the operational implementation differed by phase and local context. In early deployments, participant-selected sampling within assigned polygons was used to support feasibility under real field conditions. In later deployments, GIS-supported candidate locations, Google Maps guidance, and sequence-generated allocation were introduced to strengthen spatial guidance, improve dispersion, and reduce convenience-based clustering. Across countries and regions, implementation was adapted to local accessibility, safety, respondent availability, and institutional coordination capacity, while preserving the common requirement of geographically distributed sampling within defined spatial boundaries. Thus, protocol refinement moved from flexible participant-selected sampling toward stronger GIS-guided allocation, while retaining field-level substitution where households were unavailable or inaccessible.

All substitutions were recorded using the same digital documentation structure as primary sampling locations, including GPS coordinates, timestamp, photographs where appropriate, and sample ID. This ensured that substitution did not remove the record from the spatial-allocation and quality-control workflow.

Introduction section (before the survey begins):

Include a brief note about taking photos for context before starting the survey questions.

Photo instructions:

Do this with as few people in the photo as possible, and make sure to note the nearby landmarks, such as distinctive shops, temples, etc. that may aid in identification.



Figure S2: Photo No. 1 Street Overview (left): Take a photo of the street from about 100 meters away; Photo No. 2 House and Surroundings (center): Photograph the target house along with 2 neighboring houses; Photo No. 3 House Gate and Number (right): Ensure a clear photo of the house gate from a distance that captures 75% of the photo, showing the house number if available.

Sampling sources and households in field

Each field visit is carried out by a field team consisting of one leader and 2-5 participants. Field visits must be conducted in teams of at least two people. Solo field visits are not permitted for safety reasons.

Upon arrival at the household, the field team requests consent to conduct the survey. If the household agrees, the survey is carried out in full. If the household declines, the location is recorded in the application, and the survey is closed accordingly to document non-participation. When a household was unavailable, declined participation, or could not be accessed safely, the field team proceeded to the nearest eligible

neighboring household within the same polygon. This substitution was allowed only within the assigned polygon and was documented to preserve the intended spatial distribution while minimizing non-response-related data loss. This information is essential and considered part of the overall dataset.

If the household uses a disinfection or filtration system of any kind, two separate surveys should be conducted: one for the water after treatment (e.g., post-RO) and one for the untreated water entering the system. If there are additional water sources, such as a well or groundwater, a separate survey may also be conducted for each of them.

Field testing and documentation SOP:

Field operators performed on-site screening using portable field-testing kits according to the manufacturer's instructions and the standardized PWD field protocol. Results were entered directly into the digital survey form under the corresponding sample ID. Where values appeared inconsistent with the expected range or field context, participants were instructed to repeat the measurement when feasible and document the repeated value. Photographs of biological kits and relevant field conditions were retained for verification. These procedures were used to reduce transcription errors, support cross-checking, and improve consistency across decentralized field teams.

In the household and the lab:

1) Sample collection and ID submission:

- Each water sample collected for the survey must be accompanied by a unique identification (ID) submission to track and record the analysis results accurately.
- In the ID submission example, "Priya Minor" is recorded with a submission date of March 4th and a time of 1:05 pm IST. Her ID submission, "PM04031305," encapsulates her initials "PM," the date "0403" (for March 4th), and the time "1305" (for 1:05 pm), demonstrating a systematic approach to creating unique identifiers. This method efficiently combines personal initials, date, and time to generate a distinct and traceable ID for each submission.

• It is most important to correctly label ID submission and clearly record it with a permanent marker on the sample tube and cap. In order to prevent it from being erased, stick transparent adhesive tape over the writing.

• Ensure that each sample container is labeled with the corresponding unique ID before proceeding with the collection.

2) Entering results:

• All lab results are entered with the accompaniment of the manager of the team.

• After conducting laboratory tests, enter the results into the survey form according to the specified requirements and following the instructions provided, using survey SOP. Take a photo of all ten biological tests with ID submission and save, for example (Figure S3):



Figure S3: Biological presence/absence test results for ten household water samples after incubation, with each bag labeled by sample ID for traceability.

• Pay attention to the unit of measurement and ensure accuracy when entering data.

3) Realistic result ranges:

• The survey form includes realistic result ranges for each parameter tested. These ranges are designed to help participants identify any potential errors or anomalies in real-time.

• Review the expected result ranges before entering the data to ensure the results fall within the acceptable limits.

4) Handling exceptional results:

• If a result falls outside the realistic range (written in the caption under the parameter) and is considered exceptional, a repeat measurement must be performed to verify the accuracy of the initial result.

- Document both the initial and repeated results in the survey, noting any discrepancies and the actions taken to resolve them.

5) Survey submission:

- Upon completing all required tests and data entry, review the survey to ensure that all sections are fully and accurately completed.
- Check that all data entries match the recorded results and that no sections are left incomplete.

6) Review and verification:

- Before submitting the survey, perform a final review to verify the completeness and accuracy of all information.
- Submit the completed survey as per the prescribed method, ensuring it is received by the system for processing and analysis.
- Group leaders or designated reviewers should spatially guide select surveys for detailed checking, comparing entered data against original sample IDs and recorded test results to confirm consistency and correctness.
- Any discrepancies found during verification should be addressed immediately, with corrective actions taken and documented to maintain the integrity of the survey process.

Cross-verification methods (used and prioritized)

1) Task duplication (redundancy / cross-assignment)

- Two participants survey the *same household* independently.
- Conducted on the same day or with a short gap.
- Participants are unaware of the duplication to prevent bias.
- Results are compared for consistency.

2) Train the trainers (TTT) - alumni engagement

- Alumni conduct silent surveys during participant visits.
- They complete the form independently without interaction.

- Useful for observation, training, and quality assurance.

3) Shadow verification (future option)

- Participant's surveys while another silently observes and records method quality.
- Focuses on process integrity rather than data duplication.

4) Expert validation by interns and train the trainers

- Revisit 10% of households after their own fieldwork.
- They repeat the survey and water collection.
- Used sparingly due to logistical complexity.

Method:

- The coordinator using an Excel table with the submission ID, latitude, longitude, along with the household head's name, gender, and age. Sharing with the participant's number of locations based on the number of participants, 10 for each.
- For the participant's convenience, the participant opens a "new list" in Google Maps in advance. The participant adds all the coordinates and marks them on the map to create a pre-planned travel route. This way, they can identify any unusual location or one that is far from the others, report it to the coordinator, and request a change if needed.
- The participant should ask and confirm whether a survey was conducted about a month or two ago and mention it. "Do you remember it?" If not, ask, "Does this name sound familiar to you? Perhaps it's your wife or another household member." This should be done with utmost sensitivity, providing reliable information about the project, its purpose, and objectives. The participant should explain the significance of returning to the household and the importance of this in a pre-prepared written paragraph:

“We are from People’s Water Data, collecting a second water sample in your household is important for monitoring how stable the water quality remains over time. This will help identify any changes that may have occurred in your specific location and allow for a better understanding of the factors

influencing water quality. Regular water sampling enables more accurate recommendations for improving water purification in your area in the future and making sure that you are drinking a safe water”.

- If this is not the correct household, proceed with the survey and the entire process by opening a new ID according to all the guidelines.
- And if it is the correct household and they request to see results from a previous visit, provide them with a QR code containing the ID.
- If it is correct household provide a new ID related to the old one, and conduct a full survey.

Cross-referencing between IDs:

- In the ID entry tab, a new ID will be entered alongside the original old ID.
- The issue needs to be checked in the survey software, as the ID entry tab is limited to a certain number of characters.
- It is also necessary to check whether the results can be analyzed and how this can be entered into the platform automatically in a computerized manner.

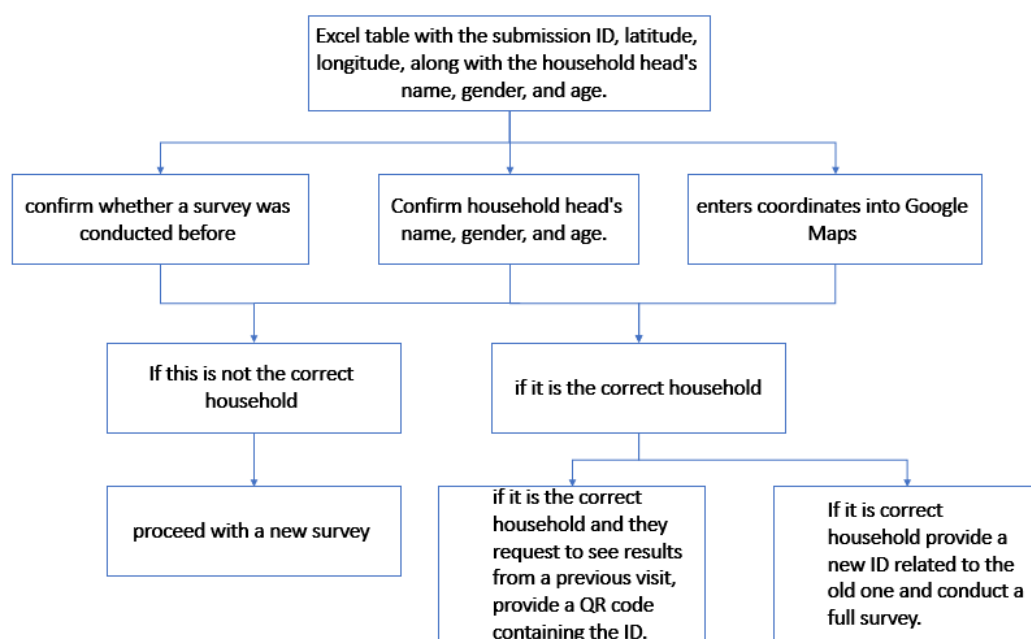


Figure S4: Household Verification and Follow-Up Survey Workflow

396

397 **Quality control measures (future plan):**

398 1. **Embedded control questions in the survey** - questions such as “What color is the tap?” at the
399 beginning and a similar one at the end to check consistency. If there is a contradiction, it may indicate that
400 the survey was not conducted properly.

401 2. **Cross-checking between groups** - One group conducts the survey, while another group verifies the
402 data, images, and locations – without visiting the field.

403 3. **Pre-planned control samples** - adding samples with known values in advance – to check the accuracy
404 of the testing process.

405 4. **Expert validation** - as we already do, after participants have entered the results and completed the tests.
406 At a regular frequency – for example, 10% of all samples.

407 **Responsibilities:**

- 408 • Managing level responsible follow up on data tracking, perform data analysis for sharing on the
409 People’s Water Data platform, and share the data with lab center coordinators.
- 410 • Lab center coordinators (Lab center level) must ensure that participants are trained on the SOP and
411 comply with its guidelines during the survey process.
- 412 • Group leaders (ground level) will spatially guide select and conduct a thorough review and verification
413 of two surveys for each participant: the first survey and one other from the remaining nine and will
414 submit that in the survey form.
- 415 • Participants (base level) are responsible for collecting samples, performing tests, recording results, and
416 submitting the completed surveys.

417 **Documentation:**

418 Maintain records of all surveys, including sample IDs, test results, any anomalies, and corrective actions
419 taken. These records should be stored securely and accessible for review or audit purposes.

420 **Review and update:**

This SOP should be reviewed on-going and updated as necessary to reflect changes in procedures or field requirements.

4. Ethical committee approval

This study was conducted following ethical approval from the Institutional Ethics Committee (IEC) of the Indian Institute of Technology Madras (Approval No. IEC/2024-02/PT/16, valid from 21 June 2024 to 20 June 2027). The research titled “A Hybrid Course on Water Quality - An Approach to People’s Water Data” involved participants and community respondents across multiple regions. All participants provided informed consent prior to participation. Field operators completed digital consent documentation through the NPTEL platform, and household respondents provided voluntary informed consent prior to survey administration. The study adhered to the Indian Council for Medical Research’s (ICMR) National Ethical Guidelines for Biomedical and Health Research Involving Human Participants (2017). Personal identifiers were removed prior to analysis, and data were stored using secure, access-controlled systems to ensure confidentiality and regulatory compliance.

Household respondents were informed that participation was voluntary and that refusal would not result in any penalty or loss of benefit. When a household declined participation, the refusal was documented only for field-completion purposes, and the team proceeded according to the substitution protocol within the same polygon.

Photographs were collected for field verification and spatial validation purposes. Field teams were instructed to minimize the inclusion of identifiable individuals in photographs whenever possible.

5. Laboratory analysis and measurements

Table S4 in the Supporting Information provides a comprehensive summary of the water quality testing parameters, methods, and standards used in this study. It includes specific criteria from both the WHO and the BIS, ensuring alignment with international and local standards.^{1,22,27} Key parameters such as turbidity,

pH, temperature, total dissolved solids (TDS), and indicators of microbial presence (e.g., *E. coli*) are detailed along with their acceptable and permissible limits. Additional notes briefly describe why each parameter is important. For example, turbidity can indicate possible microbial contamination, and pH affects chemical stability and treatment effectiveness. **Table S4** outlines the regulatory and analytical framework that guides parameter selection and testing protocols within the decentralized monitoring system.

Table S4. Overview of the water quality parameters, methods, and standards of the WHO and BIS.^{1,22,27}

Parameters	Typical Instrument Resolution	Standards WHO and BIS.	Notes
Turbidity	0.1–0.01 NTU	WHO: ≤ 5 NTU Ideally <1 NTU BIS: AL ≤ 1 NTU, PL ≤ 5 NTU AL - Acceptable limit PL - Permissible limit	Indicator of microbial presence, and particles from catchment to point of use
Oxidation-Reduction Potential (ORP)	± 1 mV	This value has to be determined on a case-by-case basis; universal values cannot be recommended (WHO)	
pH	± 0.01	6.5-8.5	Affects corrosion and quality
Temperature		WHO: Wherever possible, $25^{\circ}\text{C} > \text{Temp.} > 50^{\circ}\text{C}$. Preferably:	Influences density, solubility, DO in water and reaction rates. Rising temperatures support microbial growth

	$\pm 0.1\text{ }^{\circ}\text{C}$	$20\text{ }^{\circ}\text{C} > \text{Temp.} > 50\text{ }^{\circ}\text{C}$	
Total Dissolved Solids (TDS)	$\pm 2\% \pm 1$ or mg/L	WHO: $\leq 600\text{ mg/L}$ BIS AL: $\leq 500\text{ mg/L}$, PL $\leq 2000\text{ mg/L}$	TDS becomes unpalatable at levels $> 1000\text{ mg/l}$ (WHO)
Conductivity	$\pm 1-2\%$ or $1 \pm$ $\mu\text{S/cm}$	No alternative: $> 2500\text{ }\mu\text{S/cm}$ but not recommended	Indicates potential pollution
Free/Total Chlorine, Hardness, Alkalinity	NA	WHO: Free chlorine $\leq 5\text{ mg/L}$ Hardness (calcium ions, taste threshold) $100-300\text{ mg/L}$. BIS: Free residual chlorine: AL ≤ 0.2 , PL ≤ 1.5 . Alkalinity and hardness as CaCO_3 : AL ≤ 200 , PL $\leq 600\text{ mg/L}$	Excessive hardness may elevate health risks; chlorine $> 4\text{ mg/L}$ can impact breathing. Hardness: soft ($0-60\text{ mg/L}$), moderate ($60-120\text{ mg/L}$), hard ($120-180\text{ mg/L}$). Above $\sim 200\text{ mg/l}$ may cause scale deposition in water system
Dissolved Oxygen (DO)	$\pm 0.1\text{ mg/L}$ or $\pm 2\%$	WHO: No health-based guideline value is recommended.	Low levels lead to mineral dissolution, affecting taste and ecosystem health. Very high levels of DO may impair corrosion of metal pipes
Total coliforms, <i>E. coli</i>	NA	WHO: Total coliforms $0/100\text{ mL}$ <i>E. coli</i> $0/100\text{ mL}$. Should not be present in any 100 mL sample	Indicator of fecal contamination. <i>E. coli</i> provides conclusive evidence of recent fecal pollution

		BIS: Shall not be detectable in any 100 mL sample	
--	--	--	--

455

456 The following equipment and methods were used (Table S5):

457 1) Thermo Scientific multispectral sensor, Catalog No. ECPCWP45004K is used for measuring pH,
458 Oxidation-Reduction Potential (ORP), Conductivity, Total Dissolved Solids (TDS), and Temperature.

459 2) Thermo Scientific turbidity sensor, Catalog No. ECTN100IR

460 3) YSI's Dissolved Oxygen (DO) sensor, Catalog No. ECOSENSE ODO200

461 4) Hach 5 IN 1 dip kit (2755250 Multi Parameter test strips) for testing Free chlorine, Total chlorine, Total
462 hardness, Total alkalinity, pH.

463 5) Aquagenx's biological testing kits for *E. coli* and total coliforms in 100 ml volume, according to catalog
464 No. CBT-ECTC-PA-100. All parameters are measured on-site. At the last session, the pilot of the HiMedia
465 test strip K016 was used (<https://www.himedialabs.com/in/k023-hiselective-e-coli-test-kit.html>) for
466 bacteriological analysis.

467 Hach Test Strips provide information on five chemical and physical parameters: Total chlorine (0-10
468 mg/L), free chlorine (0-10 mg/L), total hardness (as CaCO₃, 0-425 mg/L), total alkalinity (as CaCO₃, 0-
469 240 mg/L), and pH (6.2 - 8.4). To conduct a test, the strip was immersed in the water sample for 30 seconds.

470 The color on the strip was then compared to the reference block on the container to determine the results.

471 AquagenX *E. coli* potable water quality field test kits (Presence/Absence, P/A) were used to test water
472 samples for *E. coli*. For the P/A tests, 100 mL of the tested water was carefully poured into the field kit
473 bag. After a sodium thiosulfate pill dissolved, bacterial broth powder was added and dissolved in the kit
474 bag. The bags were sealed and incubated in an incubator at 37 °C per the manufacturer's protocol for 24
475 h. Post-incubation, results were observed in ambient light, with yellow/yellow-brown indicating a negative
476 result (absence) and blue/blue-green indicating a positive result (presence).

477 **Table S5** Water quality parameters tested onsite and offsite ¹⁴.

Parameter	Location	Class	Test details
ORP, mV	Offsite	Physical	Portable multiparameter probe (Thermo scientific ECPCWP45004K)
Conductivity, $\mu\text{S}/\text{cm}$	Offsite	Physical	Portable multiparameter probe (Thermo scientific ECPCWP45004K)
TDS, mg/L	Offsite	Physical	Portable multiparameter probe (Thermo scientific ECPCWP45004K)
pH	Offsite	chemical	Portable multiparameter probe (Thermo scientific ECPCWP45004K)
Temperature, $^{\circ}\text{C}$	Offsite	Physical	Portable multiparameter probe (Thermo scientific ECPCWP45004K)
Turbidity, NTU	Offsite	Physical	Portable multiparameter probe (Thermo scientific ECTN100IR)
DO, mg/L	Onsite	Physical	Portable multiparameter probe (YSI ECOSENSE ODO200)
Coliform, P/A 100 mL	Onsite	biological	Aquagenx CBT-ECTC-PA
<i>E. coli</i> , P/A 100 mL	Onsite	biological	Aquagenx CBT-ECTC-PA
Free chlorine, mg/L as CaCO_3	Onsite	chemical	Hach CAT. NO. 2755250
Total chlorine, mg/L as CaCO_3	Onsite	chemical	Hach CAT. NO. 2755250
Total hardness, mg/L as CaCO_3	Onsite	chemical	Hach CAT. NO. 2755250
Total alkalinity, mg/L as CaCO_3	Onsite	chemical	Hach CAT. NO. 2755250
pH	Onsite	chemical	Hach CAT. NO. 2755250

478 Notes: ORP, oxidation-reduction potential; TDS, total dissolved solids; DO, dissolved oxygen; P/A, presence/absence.

479
480 Field kits, probes, and laboratory instruments were not permanently installed monitoring sensors. They
481 were portable screening or analytical tools used during field visits or laboratory-supported measurements.

482 Therefore, the term “sensor” in this study refers only to portable measurement instruments and not to fixed
483 in situ monitoring infrastructure.

484

6. Health survey tool

To complement physicochemical water quality measurements with structured contextual data, the PWD framework implemented a 430-response “Pathuvampalli Family Health Survey Tool” to document environmental, behavioral, and health-related variables relevant to household water conditions. The survey included standardized sections on demographic characteristics, water use and treatment practices, sanitation infrastructure, surrounding land use, and reported household health conditions. Each module was uniformly applied across all surveyed households to ensure consistency, comparability, and reduced data redundancy. The survey was administered and quality-controlled by the designated field coordination team in Coimbatore, who oversaw implementation, data management, and verification procedures. Further details are provided in **Table S6**.

Socio-health perspectives on household water security

Along with testing water quality parameters, structured household surveys were conducted to document water-use practices and self-reported health indicators relevant to household water security. These issues highlight the need to extend drinking water quality control beyond the point of distribution to the point of consumption.¹⁵ Research indicates that improved water sources, better hygiene practices such as hand washing and sanitation, and household-level water treatment significantly reduce diarrheal diseases in developing countries.¹⁴ However, previous field studies identified household practices that contribute to contamination risks showed that in rural settings across multiple African countries, the use of wide-mouth storage containers, water transfer between containers, and failure to cover storage vessels were all significantly associated with elevated *E. coli* contamination in household drinking water,¹⁵ and dipping handheld utensils into water instead of pouring or using a tap.¹⁵

To systematically document these variables, the PWD survey tool incorporated structured health and hygiene modules (**Table S6**). In Coimbatore, respondents reported on waterborne illnesses, hygiene

508 behaviors, and indirect economic impacts such as medical expenditure and lost workdays. A total of 430
509 household health-practice surveys were completed in this region. These behavioral and health indicators
510 were subsequently analyzed alongside objective water quality measurements to examine associations
511 between user practices and contamination patterns in the decentralized monitoring dataset.¹⁶

512 **Table S6.** Structure of the Pathuvampalli family health survey tool.

Section	Variable Group / Question Type	Description / Example Items	Response Type
General Information	Date, Time, Consent, GPS location, House No., Street, Ward	Auto-filled metadata fields ensuring traceability and ethical compliance	Auto-set / Yes–No / Numeric
Household Characteristics	Type of house, presence of well or tank, water source, farmland proximity	Identifies environmental exposure and infrastructure	Multiple choice
Water Use and Treatment	Source of drinking water, treatment method (RO, boiling, none), storage practices	Captures household-level water management and treatment behavior	Multiple choice / Text
Agricultural Context	Nearby crops, fertilizer/pesticide use	Links agricultural practices to potential contamination	Checkbox / Text
Household Sanitation	Toilet and septic tank presence, greywater disposal	Evaluates sanitation and waste management	Multiple choice
Demographics	Adults (1–6) and children (<18 years): name, age, sex, education, occupation, relation	Records family structure and socio-economic background	Structured fields
Lifestyle & Health (per adult)	Alcohol, smoking, and tobacco use	Identifies behavioral risk factors	Multiple choice

Medical History (per adult)	Diabetes, hypertension, kidney/liver/cardiac disease, cancer, COVID-19, others	Captures non-communicable and infectious disease prevalence	Yes–No / Text
Children’s Health	Age, birth weight, illness frequency (fever, diarrhea), birth defects	Tracks early-life exposures and health indicators	Numeric / Multiple choice
Mortality Record	Deaths in last 10 years, cause (disease, accident, aging)	Assesses long-term household health trends	Multiple choice / Text
Water Sampling Metadata	Drinking and source sample IDs, remarks	Links survey data to physical samples and lab analysis	Text (coded IDs)

7. Household feedback

To ensure transparency and community engagement, the PWD framework developed a standardized reporting mechanism that links monitoring outputs to household-level notifications. The reporting framework combines EpiCollect5 for cellular phone-based data entry and GIS mapping for spatial visualization, allowing real-time tracking of water quality across surveyed households.

Each participating household received a personalized digital water quality report, summarizing measured parameters, their compliance with WHO, BIS, and EPA standards, and color-coded indications of concern (e.g., red for non-compliance). Reports were generated through automated templates and shared via secure digital links, ensuring that even participants in remote areas could access their results.

The reports were written in clear, non-technical language and included explanations for each parameter (such as pH, turbidity, hardness, alkalinity, free and total chlorine, total dissolved solids, and bacterial presence). Where contamination was detected, standardized advisory notes were provided, including recommendations for improved storage practices, point-of-use disinfection, and periodic maintenance of household RO systems. These recommendations were aligned with the Jal Jeevan Mission and WHO Guidelines for Drinking-Water Quality to maintain consistency with national and international standards.

To complete the reporting cycle, households were invited to submit comments or follow-up updates through an integrated online form accessible via QR code on each report. This structured response channel enabled documentation of household-level reactions to reported findings without constituting a formal intervention program. An example of the standardized household report format is detailed in **Text S7** (Water Test Report).

Household feedback and recommendations for improved practices

A subset of participating households (10% of the total sample) received a personalized water quality report aligned with WHO standards. Reports were delivered digitally and formatted in clear, non-technical language to facilitate comprehension. When measured values exceeded reference standards, standardized advisory notes were provided in accordance with predefined guidance protocols. These included recommendations on storage practices, filtration, disinfection, and maintenance procedures, where applicable. Households were invited to provide optional feedback through an online form to document response patterns and user perception (see **Table S7**). This feedback mechanism enabled documentation of how information was shared with households and how they responded, without constituting a formal intervention framework.

Water Test Report

Sample ID: Date of Collection: Time of Collection:

Name & Address:

Thank you for participating in the [People's Water Data](#) initiative.

This is a summary of the water quality test results for the sample collected by our volunteer from the specified location and date. These results represent the water quality at the time of testing. The table below provides the parameters tested, their measured values, and the corresponding reference limits based on WHO standards.

Table S7. Water test report

Parameter	Measured Value	Reference Limit	Compliance
pH Level	[Measured Value]	6.5 - 8.5	[Compliant/Non-Compliant]
Turbidity (NTU)	[Measured Value]	≤ 5.0 NTU	[Compliant/Non-Compliant]
Total chlorine	[Measured Value]	≤ 1.5 mg/L	[Compliant/Non-Compliant]
Free chlorine	[Measured Value]	≤ 1.5 mg/L	[Compliant/Non-Compliant]
Hardness	[Measured Value]	≤ 600 mg/L	[Compliant/Non-Compliant]
Alkalinity	[Measured Value]	≤ 600 mg/L	[Compliant/Non-Compliant]
Total Dissolved Solids (TDS)	[Measured Value]	≤ 2000 mg/L	[Compliant/Non-Compliant]
ORP	[Measured Value]	≥ 200 Mv	[Compliant/Non-Compliant]
Conductivity	[Measured Value]	≤ 1500 μS/cm	[Compliant/Non-Compliant]
Total Coliform Bacteria	P/A	Absent	[Compliant/Non-Compliant]
<i>E. coli</i> Bacteria	P/A	Absent	[Compliant/Non-Compliant]

Note: "Compliant" means the tested value is within the Permissible Reference limit, while "Non-Compliant" indicates that the tested value exceeds the Permissible limit. N/A is non-available.

The values highlighted in red are not within the reference limit.

Disclaimer: Measured values may vary due to inherent variability or external factors.

For more details:

1. Bureau of Indian Standards. (2012). *IS 10500: 2012 – Drinking water specifications*. Bureau of Indian Standards. <https://law.resource.org/pub/in/bis/S06/is.10500.2012.pdf>
2. World Health Organization. (2017). *Guidelines for drinking-water quality: Fourth edition incorporating the first addendum*. <https://www.who.int/publications/i/item/9789241548151>
3. U.S. Environmental Protection Agency. (2021). *National primary drinking water regulations*. <https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations>
4. Jal Jeevan Mission, Government of India. <https://jaljeevanmission.gov.in/>

5. A Hybrid Course on Water Quality – An Approach to People’s Water Data. <https://elearn.nptel.ac.in/shop/iit-workshops/completed/a-hybrid-course-on-water-quality-an-approach-to-peoples-water-data-2/?v=c86ee0d9d7ed>

8. Implementation scale and household-level monitoring density

Figure S5 illustrates the geographic distribution of household-level water-quality assessments conducted within the PWD framework across participating regions. Each data point represents a surveyed location. Most surveys were conducted in India, with additional sites in southern Israel and Uganda. These deployments demonstrate workflow implementation across different institutional and operational contexts, but should not be interpreted as nationally representative country-level comparisons.

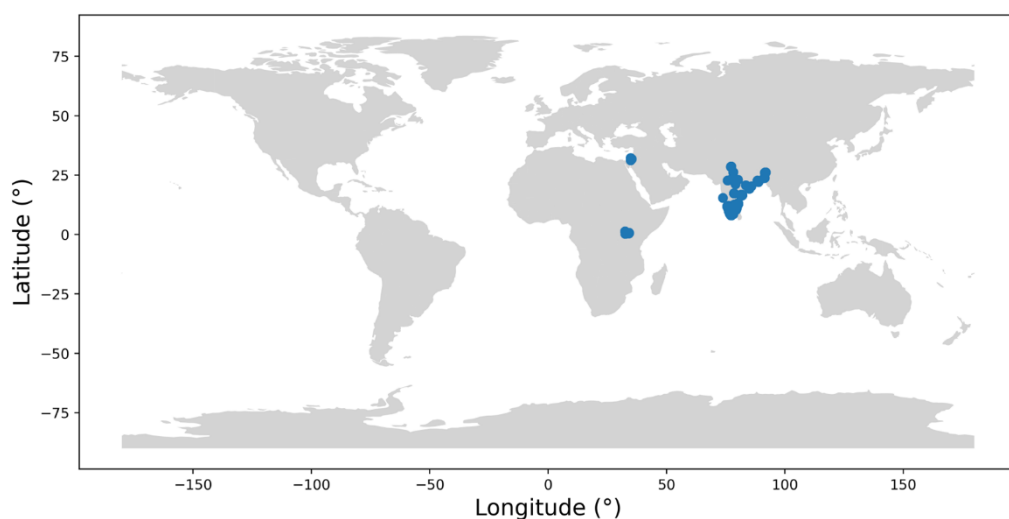


Figure S5: Geographic distribution of household-level water-quality assessments conducted within the PWD framework across participating regions. Most surveys were conducted in India, with additional sites in southern Israel and Uganda.

9. Variability in survey duration

Figure S6 complements the analysis presented in Figure 5 by examining average response times across grouped question blocks rather than individual items. While Figure 5 shows the mean duration per question based on 93,204 responses from 1,194 surveys, Figure S6 aggregates 3,900 observations from 650 surveys into six conceptual categories. This block-based approach provides a broader view of time allocation patterns. The comparison between the two datasets indicates that a clearly structured survey architecture

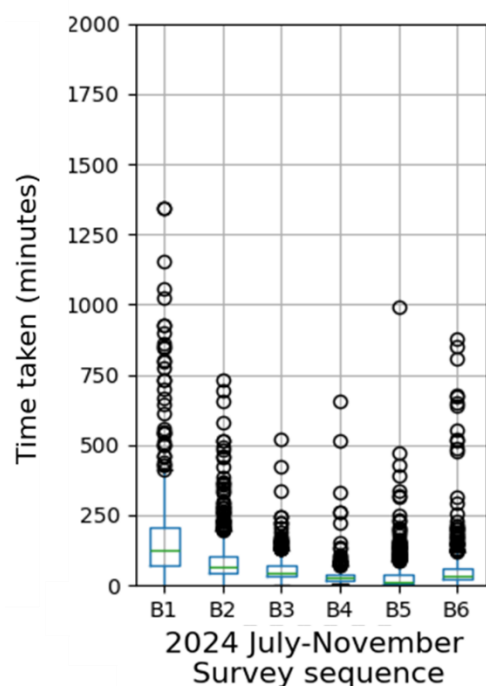


Figure S6: Survey completion time distribution grouped into six main question blocks, reflecting differences in engagement and data-entry behavior (n = 650).

10. Agreement rates

To evaluate the reliability and internal consistency of household survey responses, agreement rates were calculated for all major questions by comparing field-collected responses with expert-verified records.

Agreement rate was calculated as the number of matched participant–expert records with the same response category, presence/absence result, or field-test category divided by the total number of matched records for that variable, multiplied by 100. Only matched records with sufficient information for comparison were included in the denominator. Records with missing values, incomplete identifiers, or unmatched household/sample IDs were excluded from variable-specific agreement calculations.

Because the PWD validation dataset included a mixture of categorical survey responses, presence/absence microbial results, and semi-quantitative field-strip categories, agreement percentages were used as the primary harmonized validation metric across data types. In previous community-based water-quality validation work, different metrics were applied according to parameter type: categorical strip-test results were evaluated using category agreement, microbial presence/absence results were evaluated using confusion-matrix-derived measures such as accuracy, precision, and sensitivity, continuous sensor readings were evaluated using relative differences and coefficients of determination, and categorical survey responses were evaluated using agreement rates and, where appropriate, Cohen's kappa. In the present PWD analysis, agreement percentages were used as the common metric across survey responses, presence/absence microbial results, and semi-quantitative field-test categories to allow consistent comparison across heterogeneous data types. More complex statistics, such as Cohen's kappa, confusion matrices, or continuous-error metrics, may be appropriate for future parameter-specific validation once larger paired datasets are available for each measurement type.

For semi-quantitative strip-based parameters, agreement can be assessed at different levels of strictness. Exact agreement requires participant and expert readings to fall in the same color category. A less strict category-based agreement may allow adjacent color categories, reflecting the semi-quantitative nature of color-block strip tests and their limited visual resolution. This approach follows previous community-based water-monitoring validation work using Hach test strips, where both exact and category-tolerant agreement were reported. In the present PWD analysis, the agreement values reported in Tables S8 and S9 are based on matched categories unless otherwise specified.

This assessment provides a quantitative measure of data consistency within the decentralized, quality-controlled monitoring framework of PWD. The expert follow-up dataset included 307 revisited households; however, only 123 records were retained for deep-dive validation after SOP-based screening, pilot refinement, household matching, and completeness checks. The remaining records were used for

operational review but were not included in the final variable-specific agreement tables when matching or completeness criteria were not met. Because follow-up verification depended on household availability, field accessibility, and feasibility of revisiting the same location, the validation subset should be interpreted as an operational reliability assessment rather than a fully spatially guided sample.

The analysis covered both observational and respondent-based questions related to water source identification, container hygiene, treatment practices, and perception of water safety. Results indicate high consistency (>80 %) for most observable parameters, such as source type, container material, and placement, while lower agreement was observed for interpretive or recall-based responses, particularly those concerning maintenance frequency or recent cleaning history. These findings support the reliability of standardized question formats and highlight areas for further refinement in survey structure and protocol clarity. Detailed agreement rates for all surveyed parameters are summarized in Table S8.

Table S8. Agreement rates for all household survey parameters between participant and expert responses.

Question	Survey Column	Agreement Rate (%)	
Where do they collect the drinking water from?	20_Where_do_they_fet	85	
Specify the type of container from which they fetch the water	23_Specify_the_type_	72.5	
Specify the type of treatment \ filter unit from which they fetch the water	60_What_kind_of_filt	9.26	
Did the respondent's hand touch the water while filling the cup?	24_Did_the_responden	72.5	
How was the water taken out of the container?	25_How_they_fetched_	82.5	
Was this drinking water ever stored in a container?	31_Was_the_drinking_	95	
What material is the container made of?	32_What_material_is_	88	
What is the size of the container in literes?	34_What_is_the_size_	71	

Is the container placed less than 2 meters away from cooking environment?	35_Is_the_container_	94	
Is the container on the ground?	36_Is_the_container_	85	
When did you last clean your container?	38_When_did_you_last	29	
Did you use any detergents for cleaning your container?	39_Did_you_use_any_d	76	
How do you feel about the safety of your drinking water today?	73_How_do_you_feel_a	62.16	
What is the risk of becoming ill by consuming your drinking water?	74_What_is_the_risk_	89.19	
Have you or another resident in this household suffered from one of the following within the last month?	75_Does_yourself_or_	90	
When was the last time this household didn't have enough drinking water?	76_What_is_the_last_	70	
When was the last time the RO maintenance service arrived?	61_When_was_the_last	66.67	
What brand and model of machine do you have?	62_Mention_the_brand	33.33	
Is anything done to maintain the filter unit?	64_Is_anything_done_	100	
What does the maintenance include?	65_What_does_the_mai	100	
After how long maintenance is required?	66_What_is_the_maint	100	
Have you maintained your filter unit \ RO on the required time?	67_Do_you_maintain_y	100	
Name of respondent	10_Name_of_responden	47.5	
Number of children under 5	17_Number_of_childre	92.5	

640

641 The comparison between expert and field-based measurements was limited to parameters analyzed directly
642 on-site using a 5-in-1 dip kit and biological test. Measurements performed later in the laboratory using
643 advanced sensors were excluded due to resource and personnel constraints. The analysis thus focused on
644 field-measurable indicators, for which both field operators and expert verifiers collected data under
645 comparable environmental conditions and standardized guidance.

Results demonstrated high consistency across physicochemical parameters, total chlorine (100%), free chlorine (100%), hardness (87%), alkalinity (84%), and pH (89%), and strong agreement in biological tests of *E. coli* (87%) and total coliforms (96%). These findings confirm the reliability of decentralized field measurements for routine surveillance applications and highlight the robustness of standardized protocols within the PWD framework. Detailed agreement rates for all evaluated parameters are summarized in Table S9.

Table S9. Agreement rates for field-measured parameters.

Category	Parameter	Agreement Rate (%)
5-in-1 HACH Physicochemical Parameters	Total Chlorine	100%
	Free Chlorine	100%
	Hardness	87%
	Alkalinity	84%
	pH	89%
Biological Parameters	<i>E. coli</i>	87%
	Total Coliforms	96%

11.Data quality risks and mitigation

The PWD framework involved decentralized field data collection by trained participants across heterogeneous field settings. Therefore, several potential data-quality risks were considered during protocol design, implementation, and post-field validation. Table S10 summarizes the main risks, their potential effects, and the mitigation measures used to reduce their impact on data reliability and interpretability.

Table S10. Data quality risks and mitigation measures in the PWD decentralized monitoring framework.

Risk	Potential effect	Mitigation measure
Convenience sampling / spatial clustering	Reduced spatial representativeness and over-sampling of easily accessible locations	Polygon assignment, GIS-supported candidate locations, Google Maps guidance, and post-field GPS review
Household non-response or inaccessibility	Missing data or location-based selection bias	Substitution with the nearest eligible neighboring household within the same polygon, with documentation of the replacement location
GPS error or inaccurate geotagging	Location mismatch between household records, samples, and validation revisits	GPS metadata, timestamps, photographic documentation, and expert spatial verification
Protocol deviation	Inconsistent survey execution, field testing, or sample handling	Standardized training, SOP-based field guidance, timestamps, photographs, group-leader review, and expert cross-verification
Transcription or data-entry error	Incorrect matching between household survey, sample ID, field measurements, and laboratory results	Unique alphanumeric sample ID, digital forms, timestamp linkage, and sample-labeling procedures
Open-ended or interpretive responses	Lower agreement between participant and expert records	Use of closed-ended, multiple-choice, and observable question formats where standardization was required
Field-kit or instrument limitations	Measurement uncertainty, especially for semi-quantitative or categorical test outputs	Field kits interpreted as screening tools; laboratory analysis used for selected

Risk	Potential effect	Mitigation measure
		parameters where higher analytical precision was required
Delayed uploads or offline data entry	Misleading survey-duration records or uncertainty about real-time field execution	Timestamp review, survey-duration analysis, and outlier screening
Observer bias	Variable interpretation of household practices, storage conditions, or respondent answers	Training, simplified observable questions, standardized response options, photographs, and expert validation
Environmental variability between visits	Differences between participant and expert measurements due to temporal variation rather than operator error	Same-location revisits where feasible; validation interpreted as operational field reliability rather than continuous analytical equivalence
Sample handling or labeling errors	Loss of traceability between field sample and laboratory result	Sterile containers, permanent sample labeling, sample ID linkage, and verification before data submission
Falsification or fabricated entries	Invalid field records and reduced trust in decentralized data	GPS coordinates, timestamps, minimum expected survey duration, photographic documentation, and review of suspicious completion patterns

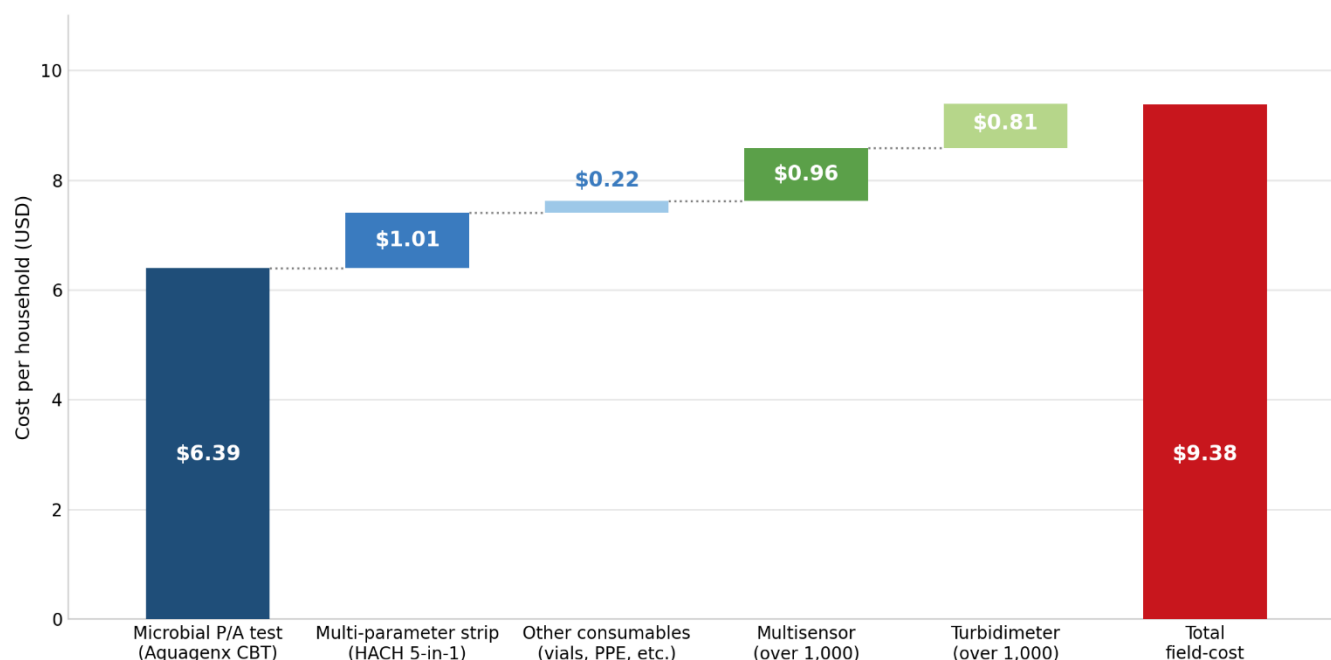
662

663

12.Resource-efficiency and monitoring-density implications

Documented direct field-cost metrics were estimated using supplier quotations and purchase records available for the PWD field workflow. These estimates are presented as documented direct equipment-and-consumables costs only, not as a full implementation cost, cost-effectiveness analysis, or comparison with certified regulatory monitoring.

The microbial presence/absence test cost was estimated at \$6.39 per household, based on \$12,781 for 2,000 tests. The multi-parameter strip cost was estimated at \$1.01 per household, based on ₹168,315.20 for 2,000 strips, converted using ₹83 = US\$1. Additional direct consumables, including vials, PPE, and related field-use items, were estimated at \$0.22 per household, giving a total documented consumable cost of \$7.62 per household. Reusable field instruments were treated as shared equipment costs. Five multi-parameter meters and five turbidity meters were purchased. Assuming use across 5,000 household assessments in total, corresponding to 1,000 household assessments per unit, the distributed reusable equipment costs were \$0.96 per household for the multi-parameter meters and \$0.81 per household for the turbidity meters.



680 **Figure S7.** Documented direct equipment and consumables cost per household assessment.

681

682 Under these assumptions, the documented direct equipment-and-consumables cost was approximately
 683 \$9.38 per household assessment, corresponding to approximately 107 household-level assessments per
 684 \$1,000. When 11 reported core water-quality parameters are considered per household, this corresponds to
 685 approximately 1,170–1,180 core water-quality parameter observations per \$1,000. These estimates exclude
 686 labor, transport, training, calibration supplies, customs and duties, laboratory validation, institutional
 687 overhead, data-platform development or maintenance, and long-term governance costs.

688

689 **Table S11.** Documented direct equipment and consumables cost assumptions for PWD monitoring-density
 690 estimates.

Component	Calculation basis	Estimated value	Interpretation
Microbial presence/absence test	\$12,781 / 2,000 tests	\$6.39 per household	Direct consumable cost for microbial presence/absence screening
Multi-parameter strip	₹168,315.20 / 2,000 strips; ₹83 = US\$1	\$1.01 per household	Direct consumable cost for strip-based physicochemical screening
Other consumables	Vials, PPE, and related field-use consumables	\$0.22 per household	Additional direct consumables included in the documented field-cost estimate
Multi-parameter meters	5 × \$964 / 5,000 household assessments	\$0.96 per household	Distributed reusable equipment cost

Component	Calculation basis	Estimated value	Interpretation
Turbidity meters	$5 \times \$807 / 5,000$ household assessments	\$0.81 per household	Distributed reusable equipment cost
Estimated direct equipment-and- consumables cost	$\$6.39 + \$1.01 + \$0.22$ $+ \$0.96 + \0.81	\$9.38 per household	Estimated direct field-cost basis used for monitoring-density calculation
Household assessments per \$1,000	$\$1,000 / \9.38	~107 households	Estimated household-level monitoring density per budget unit
Core parameter observations per \$1,000	~107 households $\times$ 11 parameters	~1,170–1,180 observations	Information-density indicator for household-level water-quality surveillance

Notes: Values represent documented direct equipment-and-consumables costs only. They include microbial presence/absence kits, multi-parameter strips, selected additional direct consumables, and distributed reusable field-equipment costs. They do not include labor, transport, training, calibration supplies, customs and duties, laboratory validation, institutional overhead, data-platform development or maintenance, or long-term governance costs. INR-denominated quotations were converted using ₹83 = US\$1 for reporting consistency. Reusable equipment costs are distributed across 5,000 household assessments, based on five purchased units of each device type and an assumed 1,000 household assessments per unit. This assumption is used only to express reusable equipment as a per-household field-cost component and should not be interpreted as a manufacturer-certified device lifetime.