Journal of Water & Health

© 2023 The Authors Journal of Water and Health Vol 21 No 12, 1834 doi: 10.2166/wh.2023.174

E. coli contamination of drinking water sources in rural and urban settings: an analysis of 38 nationally representative household surveys (2014–2021)

Thiago M. Santos [a,b,*], Andrea Wendt [a,c], Carolina V. N. Coll [a], Meghan A. Bohren [b] and Aluisio J. D. Barros [a]

[a] International Center for Equity in Health, Federal University of Pelotas, Rua Deodoro 1160, Pelotas, RS 96020-220, Brazil
[b] Gender and Women's Health Unit, Nossal Institute for Global Health, School of Population and Global Health, University of Melbourne, 207 Bouverie St, Carlton, VIC 3053, Australia
[c] Programa de Pós-Graduação em Tecnologia em Saúde, Pontifícia Universidade Católica do Paraná, Rua Imaculada Conceição 1155, Curitiba, PR 80215-901, Brazil
*Corresponding author. E-mail: tmelo@equidade.org

TMS, 0000-0002-4572-5297; AW, 0000-0002-4640-2254; CVNC, 0000-0003-0808-8230; MAB, 0000-0002-4179-4682; AJDB, 0000-0002-2022-8729

ABSTRACT

The world is not on track to achieve universal access to safely managed water by 2030, and access is substantially lower in rural areas. This Sustainable Development Goal target and many other global indicators rely on the classification of improved water sources for monitoring access. We aimed to investigate contamination in drinking water sources, comparing improved and unimproved sources in urban and rural settings. We used data from Multiple Indicator Cluster Surveys, which tested samples from the household water source and a glass of water for *Escherichia coli* contamination across 38 countries. Contamination was widespread and alarmingly high in almost all countries, settings, and water sources, with substantial inequalities between and within countries. Water contamination was found in 51.7% of households at the source and 70.8% in the glass of water. Some improved sources (e.g., protected wells and rainwater) were as likely to be contaminated as unimproved sources. Some sources, like piped water, were considerably more likely to be contaminated in rural than urban areas, while no difference was observed for others. Monitoring water contamination along with further investigation in water collection, storage, and source classification is essential and must be expanded to achieve universal access to safely managed water.

Key words: drinking water, *Escherichia coli*, global health, health inequities, water quality

HIGHLIGHTS

- Water sources classified as improved – in particular, protected wells and rainwater – had high prevalence of *E. coli* contamination.
- Piped water sources were much more likely to be contaminated in rural areas.
- Contamination was more prevalent at the point of use than at the source, but this varied significantly according to the water source.
- We provided a list of suggestions to improve monitoring of the Sustainable Development Goal 6.

INTRODUCTION

Access to safe and clean water was recognised as a human right in 2010, and the first target of the Sustainable Development Goal (SDG) 6 is to 'achieve universal and equitable access to safe and affordable drinking water for all' by 2030 (UN General Assembly 2010; United Nations 2023a). The world is not on track to reach this target, and 2 billion people still lacked access to safely managed drinking water in 2020 (WHO, UNICEF and World Bank 2022). Even though significant progress was achieved in the last 20 years, with access increasing from 62% in 2000 to 74% in 2020 – equivalent to an additional 2 billion people – regional inequalities remain alarmingly high. While 96% of the population in Europe and North America had access to safely managed water in 2020, only 30% had access in sub-Saharan Africa (WHO, UNICEF and World Bank 2022).

A safely managed water service is an improved water source, which is located on the premises, available when needed, and free from contamination (WHO, UNICEF and World Bank 2022). Improved water sources are 'those that, by nature of their design and construction, have the potential to deliver safe water' (WHO, UNICEF and World Bank 2022). For example, piped supplies and boreholes are classified as improved sources, while surface water is not (WHO 2022). Contamination refers to microbiological – typically *Escherichia coli* – and chemical contamination, including arsenic, fluoride, lead, and nitrate, depending on regional priorities (WHO, UNICEF and World Bank 2022). In case of water contamination data not being available, the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) can provide estimates for other indicators, most notably the basic water service indicator, which refers to using an improved water source with water collection time equal to or lower than 30 min (WHO and UNICEF 2021). Therefore, it is heavily reliant on the definition of an improved water source.

While the JMP can provide estimates for 211 countries (representing 99% of the global population) for the basic water service indicator, only 138 countries (representing 45% of the global population) have data for the safely managed indicator (WHO and UNICEF 2021). This limitation is because many countries and regions lack the necessary capacity and infrastructure to monitor the water contamination component of the indicator (WHO and UNICEF 2021). Nationally representative household surveys represent a valuable data source that not only can include measurements of water quality but also can allow for the disaggregation of estimates according to important dimensions of inequality and household characteristics. In particular, from 2017 onwards, the UNICEF-supported Multiple Indicator Cluster Survey (MICS) have implemented a water quality module that measures *E. coli* contamination in drinking water samples, mostly in low- and middle-income countries (LMICs) (Bain *et al.* 2021). Data are available for 38 countries.

Current evidence indicates that water contamination is high in LMICs, but it markedly varied between countries, urban and rural settings, and water sources (Bain *et al.* 2021; WHO UNICEF and World Bank 2022). A recent analysis of 27 LMICs showed that the proportion of the population using a contaminated water source ranged from 16% in Mongolia and Algeria to 90% in Sierra Leone (Bain *et al.* 2021). Furthermore, living in a rural household was associated with a 10% increase in the risk of water contamination compared to an urban household, after adjusting for water source, sanitation infrastructure, wealth, and other household characteristics (Bain *et al.* 2021). Unimproved water sources – such as surface water and unprotected wells and springs – have the highest contamination levels, but the water from all sources can be contaminated depending on the country (WHO UNICEF and World Bank 2022). Evidence on whether specific source contamination varies according to urban and rural settings is scarce. In a meta-analysis of faecal contamination in drinking water, there was only weak evidence that piped water is more contaminated in rural areas, and stratified results were not available for other sources separately (Bain *et al.* 2014).

Our objective is to investigate the risk of *E. coli* contamination in drinking water according to the water source and urban/rural setting. We aimed to examine contamination at the point of collection (directly from the source) and at the point of consumption (after storage and handling in the household). In particular, we want to compare water sources currently classified as improved or unimproved by the JMP in terms of their ability to provide water that is free from contamination in urban and rural settings.

METHODS

Data sources and study sample

We selected all MICS surveys with the water quality testing module publicly released until 8 February 2023 (UNICEF 2023). We included the most recent survey if more than one was available for the same country (Bangladesh and Nepal). Thirty-eight countries with surveys from 2014 to 2021 were included and are presented in Supplementary Table 1. MICS surveys use a multistage sample design to provide nationally representative samples for multiple indicators, including water, sanitation, and hygiene (WASH) (Khan & Hancioglu 2019). The first stage of sampling is typically based on census enumeration areas, and different allocation methods can be used. For the second stage, households in the selected enumeration areas are listed, and 20–25 households are selected using random systematic sampling (Khan & Hancioglu 2019). The sample is typically stratified by urban/rural areas and sub-national regions to provide adequate sample sizes for these areas (Bain *et al.* 2021). Our unit of analysis was the household.

E. coli as an indicator of faecal contamination

Water quality testing is performed in a subsample of the households surveyed by MICS, and the presence of *E. coli* is used as an indicator of faecal contamination, henceforth referred simply as ‘contamination’ (UNICEF and WHO 2020). Typically, 5 out of the 25 households in each survey enumeration area are selected, considering that households in the same cluster tend to have similar contamination levels since they are more likely to have similar water and sanitation infrastructure (UNICEF and WHO 2020). The field tester requests permission from a knowledgeable respondent from the household – aged 18 years older – to collect water samples: one glass of drinking water and another directly from the household's water source (UNICEF and WHO 2020). Samples of 100 ml are collected from the glass (point of use) and the source (point of collection) using a sterile Whirl-Pak collection bag or a preassembled membrane filtration apparatus (UNICEF 2016; Bain *et al.* 2021). For each sample, the field tester withdraws 1 ml of water using a disposable syringe and uses it to hydrate a CompactDry EC growth media plate (Nissui). The rest of the 100 ml sample is filtered with a 0.45-μm filter membrane (Millipore Microfil). The filtered water is discarded, and the filter is plated on the rehydrated growth media plate (UNICEF 2016; Bain *et al.* 2021). The plate is then incubated for 24–48 h in an electric incubator or in an incubator belt worn around the body of the field tester at a temperature between 25 and 40 °C (UNICEF 2016). The chromogenic enzyme (β-glucuronidase) substrate on the plate gives *E. coli* colonies a blue colour (UNICEF and WHO 2020). After incubation, the field tester counts and registers the number of blue colonies in the survey questionnaire. For quality control, blank tests are regularly performed using bottled or distilled water, usually after every 10 samples (UNICEF and WHO 2020). Water was considered free from contamination in our analyses if no *E. coli* colonies were detected and contaminated otherwise (WHO and UNICEF 2021).

Stratifiers

Water contamination was explored by the source and by the urban or rural household location. The urban/rural classification was based on each country's definition, generally dependent on the population density of the settlement, but additional criteria – such as the percentage of the population engaged in agriculture – might be used depending on regional specificities (Bartram *et al.* 2014; United Nations 2017).

The water source typically refers to where the household members collect water, not the origin of the water. If water is collected from a public tap supplied with water from a dam, the household water source would be the tap, not the dam (UNICEF and WHO 2018). In case of multiple sources, it refers to the main water source, i.e., the place from which the household members most frequently collect drinking water (UNICEF 2020). The water source is classified based on the respondent's response, whom the interviewer assists. Interviewers are trained and tested during the survey preparation and data collection phases and have access to pictorials depicting different water sources during the interview (Bartram *et al.* 2014; UNICEF 2021a).

Water source categories are fairly standardised in the most recent MICS surveys using the JMP naming convention. Improved water sources include piped to yard/plot, piped to neighbour, public tap/standpipe, tubewell/borehole, protected well, protected spring, rainwater, tanker truck, cart with small tank/drum, other forms of delivered water (such as buckets delivered at home), water kiosk, bottled water, sachet water, and other forms of packaged water (such as water coolers) (WHO and UNICEF 2021). Unimproved water sources include unprotected well, unprotected spring, and surface water (river, dam, lake, irrigation channel, etc.), and any other sources not listed (WHO and UNICEF 2021).

Protected wells are dug wells protected from runoff water and falling material, with some form of casing raised above the ground level and a top coverage structure. Protected springs are those protected by a 'spring box' around them that directs water to a pipe or cistern without contact with runoff water. Unprotected wells and springs are those that fail to meet those criteria (UNICEF and WHO 2018).

Statistical analysis

First, we calculated the percentage of households using each type of water source. We then calculated the percentage of households with contaminated water at the point of collection and the point of use according to the water source and the area of residence. Groupings with less than 25 households were dropped from this analysis due to low precision (UNICEF 2021b). All proportions were estimated with respective 95% confidence intervals (95% CIs).

We calculated the percentage of contaminated blank tests for each country as a form of quality control. Contaminated blank tests can indicate the general risk of cross-contamination in a specific survey but cannot be used to identify households where cross-contamination might have occurred.

We performed country-specific analyses and also for all countries pooled together to identify general patterns. For the pooled results, we combined all survey datasets into one and recalculated the survey weights to account for the country's population size. Thus, the pooled results are equivalent to a weighted average considering the national population as the weight. We used the 2018 population estimates (median survey year) from the World Bank website (World Bank 2022). A more comprehensive description is available in the Supplementary Material.

We used the 'survey' package (Lumley 2004) in R (version 4.2.2, R Foundation for Statistical Computing, Vienna, Austria) to take into account the complex survey design used by MICS (including weighting, strata, and clustering). We used the weights provided for the water quality testing subsample when available. Otherwise, we used the household sample weights.

RESULTS

Our sample consisted of 95,070 households in 38 countries. Using the SDG world regions, we had 16 countries in sub-Saharan Africa, five in Northern Africa and Western Asia, two in Central and Southern Asia, three in Eastern and South-Eastern Asia, five in Oceania, six in Latin America and the Caribbean, and one in Europe and Northern America (Supplementary Table 1) (United Nations 2023b).

Across all countries, tubewells or boreholes and piped water into the dwelling were the two most common sources of drinking water in the sample, which were used by 30.7% (95% CI: 29.8–31.5%) and 18.0% (17.3–18.6%) of households, respectively (Figure 1 and Supplementary Table 2). In urban settings, piped water into the dwelling was the most common source (30.5%; 29.1–31.9%), followed by tubewells/boreholes (17.5%; 16.3–18.8%). In rural settings, tubewells/boreholes were the most common source (40.5%; 39.5–41.5%), followed by unprotected wells (9.8%; 9.0–10.7%). The most common sources of drinking water varied considerably between countries: 87.4% (80.1–92.3%) of households in Tuvalu used rainwater; 85.3% (84.4–86.2%) in Bangladesh used tubewells/boreholes; 80.5% (77.5–83.3%) in Vietnam used piped water into dwelling; and 77.4% (75.1–79.6%) in the Dominican Republic used bottled water.

Across all countries, faecal water contamination – indicated by the presence of *E. coli* – was found in 51.7% (50.3–53.0%) of households at the point of collection and 70.8% (69.8–71.9%) at the point of use (Figures 2 and 3 and Supplementary Tables 3 and 4). Contamination was more prevalent in rural than in urban areas. The contamination prevalence at the point of collection was 59.5% (57.3–61.6%) vs. 39.8% (38.3–41.4%) in rural and urban areas, respectively. At the point of use, it was 82.1% (80.9–83.3%) vs. 55.8% (54.1–57.5%).

Contamination varied substantially between water sources. For households using unprotected wells, the contamination prevalence was 94.3% (92.4–95.7%) at the point of collection and 95.9% (94.0–97.2%) at the point of use. Meanwhile, for households with piped water into the dwelling, it was 27.9% (26.1–29.8%) and 37.3% (35.2–39.3%), respectively. There was a gradient in the prevalence between those sources with no clear cut-off point that could separate sources into low and high likelihood of contamination. For tubewells/boreholes, the most commonly used water source in the sample, the contamination prevalence was 44.1% (42.3–46.0%) at the point of collection and 82.4% (80.2–84.5%) at the point of use.

In general, households using unimproved sources were more likely to have contaminated water at the point of collection and use. Together with unprotected wells, both surface water and unprotected springs had a high likelihood of contamination. For surface water, the contamination prevalence was 90.1% (85.5–93.4%) at the point of collection and 93.4% (90.4–95.5%) at the point of use. For unprotected springs, it was 78.2% (71.4–83.7%) and 84.5% (79.9–88.2%), respectively.

Nevertheless, several improved sources were as likely to be contaminated as unimproved sources. In particular, protected wells and rainwater – both improved sources – were among the most contaminated water sources regardless of the setting (urban or rural). For households using rainwater, the contamination prevalence was 84.7% (81.1–87.7%) at the point of collection and 90.7% (87.9–92.9%) at the point of use. For protected wells, it was 89.0% (86.4–91.1%) and 92.1% (90.5–93.5%), respectively. This was only slightly lower than those from their counterpart, unprotected wells: 94.3% (92.4–95.7%) at the point of collection and 95.9% (94.0–97.2%) at the point of use.

Unlike wells, protection for springs was associated with strong reduction in contamination. Households using protected springs were considerably less likely to be contaminated than those using unprotected springs. At the point of collection, 31.0% (19.5–45.4%) were contaminated for protected springs and 78.2% (71.4–83.7%) for unprotected springs. At the point of use, the difference was smaller but still substantial: 63.2% (50.1–74.6%) for protected springs and 84.5% (79.9–88.2%) for unprotected springs.

There was substantial variation in contamination between piped sources. Water piped into the dwelling had the lowest prevalence of contamination: 27.9% (26.1–29.8%) at the point of collection and 37.3% (35.2–39.3%) at the point of use. Meanwhile, water piped into the compound and public taps/standpipes had a significantly higher likelihood of contamination. For water piped into the compound, the contamination prevalence was 53.5% (50.4–56.5%) at the point of collection and 68.6% (66.0–71.0%) at the point of use. For public taps/standpipes, the contamination prevalence was

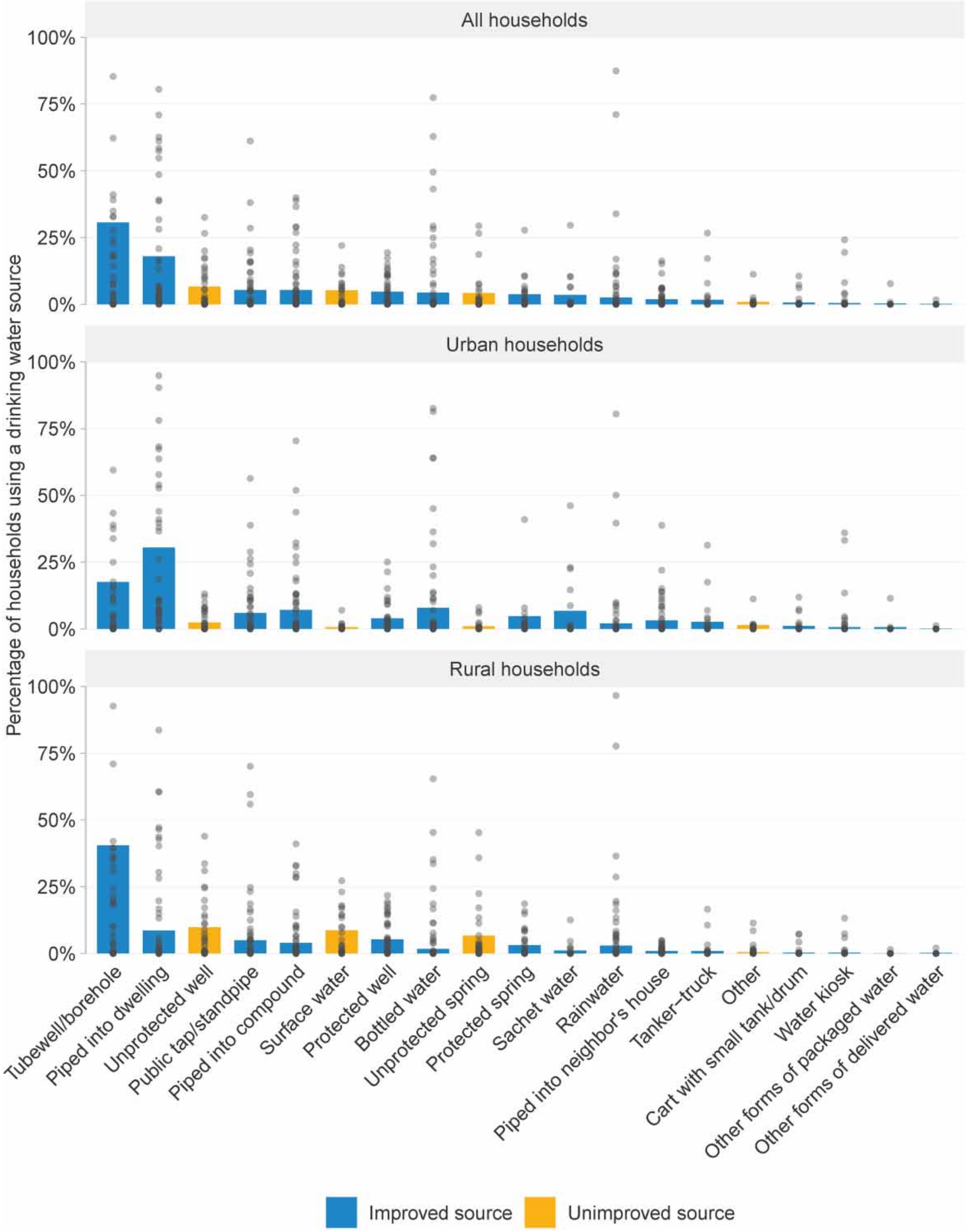

Figure 1 | Percentage of households using a specific drinking water source for all households combined (first panel) and in urban and rural settings (second and third panels, respectively). National estimates are represented by circles, while pooled estimates are represented by the bars. Water sources are ordered according to the pooled percentage of all households using those water sources.

50.1% (45.2–55.0%) and 76.3% (73.4–79.0%). Water piped into the neighbour's house had an intermediate result: 38.3% (33.0–43.9%) at the point of collection and 64.2% (57.6–70.2%) at the point of use.

The contamination prevalence of delivered water sources and water kiosks was around the 50% mark. For carts with a small tank/drum, the contamination prevalence was 49.6% (43.8–55.5%) at the point of collection and 69.0% (65.0–72.8%) at the point of use. For tanker-trucks, it was 42.0% (37.3–46.9%) and 53.2% (49.2–57.2%), respectively. For water kiosks, it was 43.7% (37.8–49.8%) and 64.8% (60.4–69.0%).

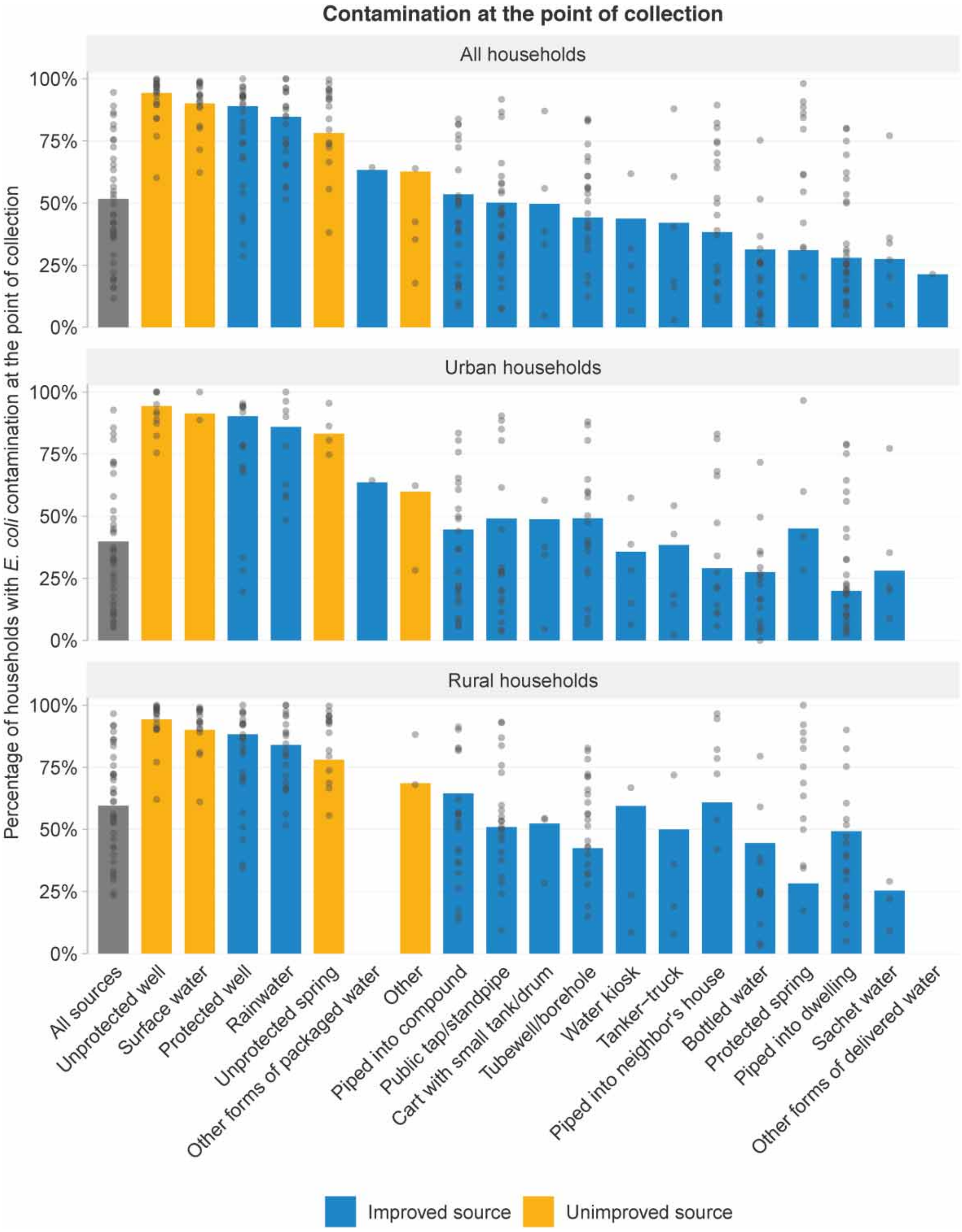

Figure 2 | Percentage of households with *E. coli* contamination at the point of collection according to the water source. Results are presented for all households combined (top panel) and in urban and rural settings (middle and bottom panels, respectively). National estimates are represented by circles, while pooled estimates are represented by bars. Water sources are ordered according to the pooled percentage of all households with *E. coli* contamination.

Packaged water sources were among the sources with the lowest prevalence of contamination. For bottled water, the contamination prevalence was 31.2% (28.9–33.6%) at the point of collection and 46.9% (45.0–48.8%) at the point of use. For sachet water, it was 27.4% (21.7–34.0%) and 44.4% (39.4–49.4%), respectively.

Other packaged and other delivered sources were only present in a few countries and had large confidence intervals, which is expected given the small sample size and the heterogeneous nature of the categories. For other packaged water sources, the contamination prevalence was 63.3% (47.3–76.8%) at the point of collection and 56.4% (43.3–68.7%) at the point of use. For

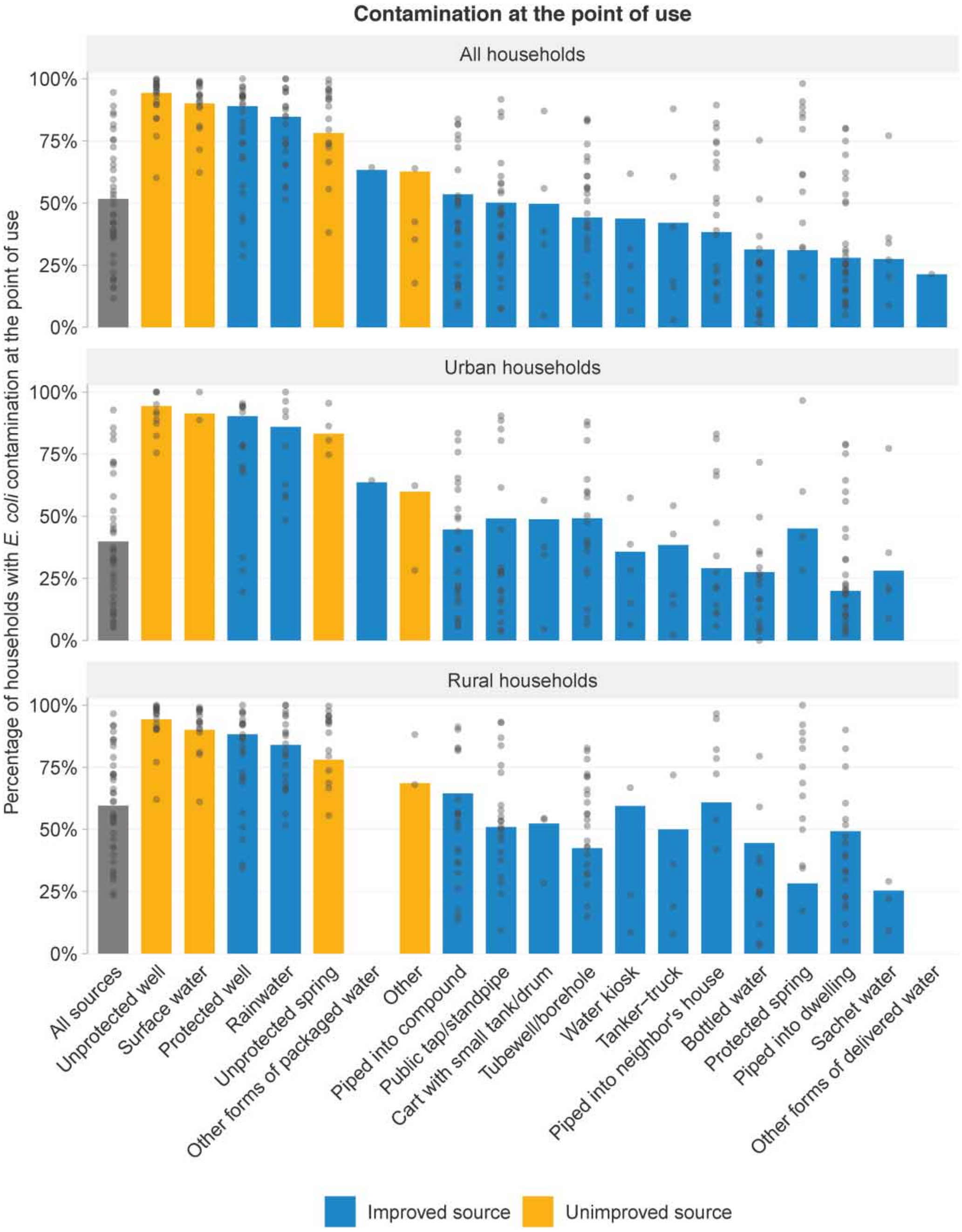

Figure 3 | Percentage of households with *E. coli* contamination at the point of use according to the water source. Results are presented for all households combined (top panel) and in urban and rural settings (middle and bottom panels, respectively). National estimates are represented by circles, while pooled estimates are represented by bars. Water sources are ordered according to the pooled percentage of all households with *E. coli* contamination.

other delivered water sources, it was 21.3% (2.2–76.6%) and 38.1% (5.7–86.3%), respectively. For the category that combined any other sources not listed, it was 62.7% (54.0–70.6%) and 80.2% (73.9–85.2%).

For all water sources, the frequency of contamination increased between the point of collection and the point of use, with the exception of other forms of package water, which had a non-significant decrease (Figure 4 and Supplementary Tables 3 and 4). Tubewells/boreholes – the most common water source in the sample – had the largest difference in the frequency of contamination: 44.1% (42.3–46.0%) at the point of collection and 82.4% (80.2–84.5%) at the point of use, or a difference of

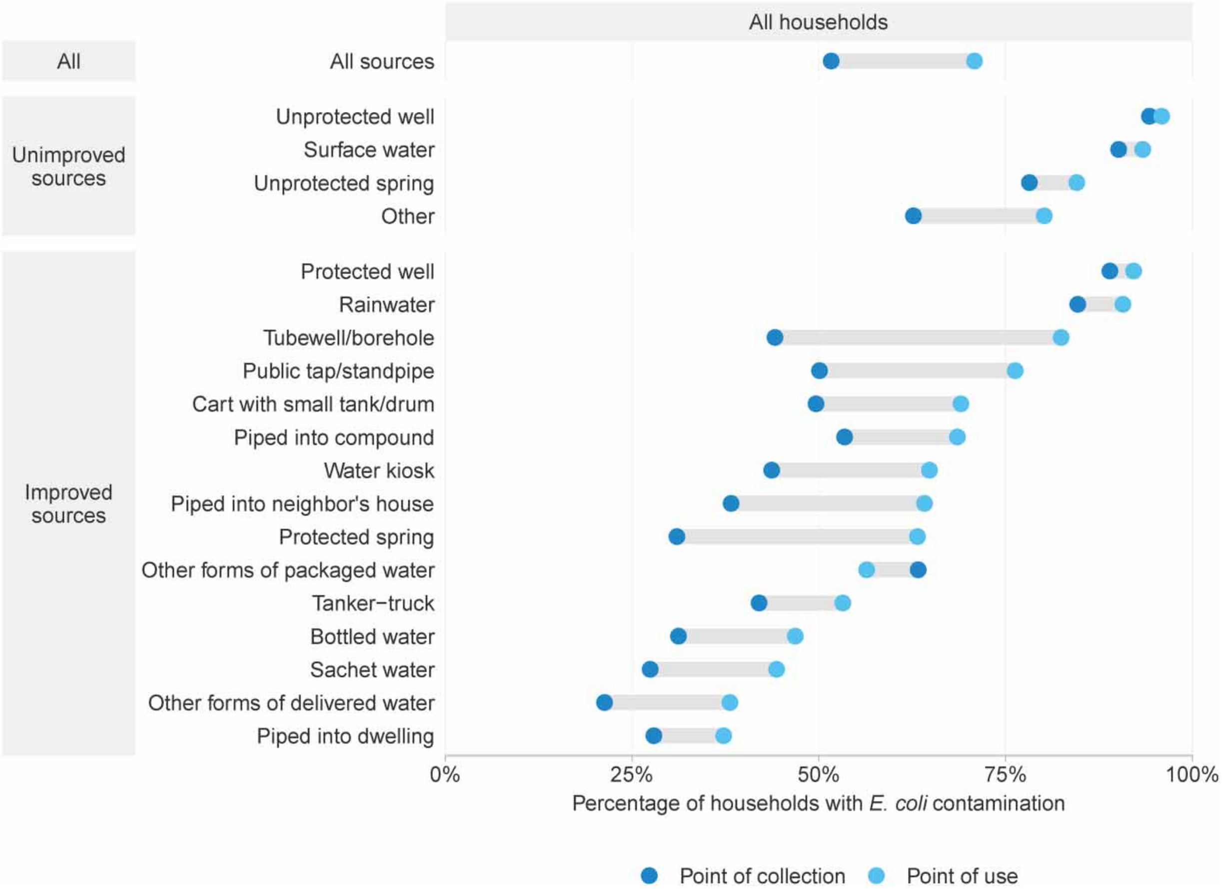

Figure 4 | Comparison in the percentage of households with *E. coli* contamination between the points of collection and use.

almost 40 percentage points. Protected springs, public taps/standpipes, piped water into the neighbour's house, and water kiosks had a difference in the frequency of contamination larger than 20 percentage points.

Differences in contamination between urban and rural households were highly dependent on water sources (Figure 5 and Supplementary Tables 3 and 4). Piped water sources (into dwelling, into the compound, and into the neighbour's house) and water kiosks had the largest absolute differences. In particular, for piped water into dwelling, at the point of collection, contamination was found in 19.9% (18.1–21.8%) of urban households and in 49.3% (44.5–54.1%) of rural households. At the point of use, contamination was found in 28.8% (26.7–31.1%) of urban households and in 59.7% (55.0–64.2%) of rural households. For many other sources, whether they had a high likelihood of contamination at the source, such as wells; an intermediate likelihood, such as tubewells/boreholes; or a lower likelihood, such as sachet water, there was no statistically significant difference in contamination at the point of collection.

There was a wide variation in the prevalence of contamination at the national level (Supplementary Tables 3 and 4). At the point of collection, contamination varied from 11.6% (4.9–25.1%) in Turks and Caicos to 94.6% (89.6–97.2%) in Tuvalu. At the point of use, it varied from 18.1% (15.8–20.6%) in Mongolia to 99.1% (98.5–99.5%) in Chad. For most countries, differences in contamination between water sources were similarly variable. At the point of collection in Madagascar, contamination varied from 8.1% (3.3–18.8%) for piped water into dwelling to 99.6% (98.1–99.9%) for unprotected wells. By comparison, at the point of use in Chad, all sources had contamination higher than 95%.

The prevalence of contamination in blank testing was low in the overall sample, 1.5% (1.1–1.8%) (Supplementary Table 5). Only three countries had a prevalence of contamination in blank testing higher than 5%: Turks and Caicos with 5.4% (1.1–23.1%), Côte d'Ivoire with 7.9% (5.4–11.4%), and Gambia with 8.8% (5.4–14.2%). There was no statistically significant difference in the prevalence of contamination in blank testing between the urban and rural settings: 1.1% (0.8–1.5%) and 1.7% (1.2–2.4%), respectively.

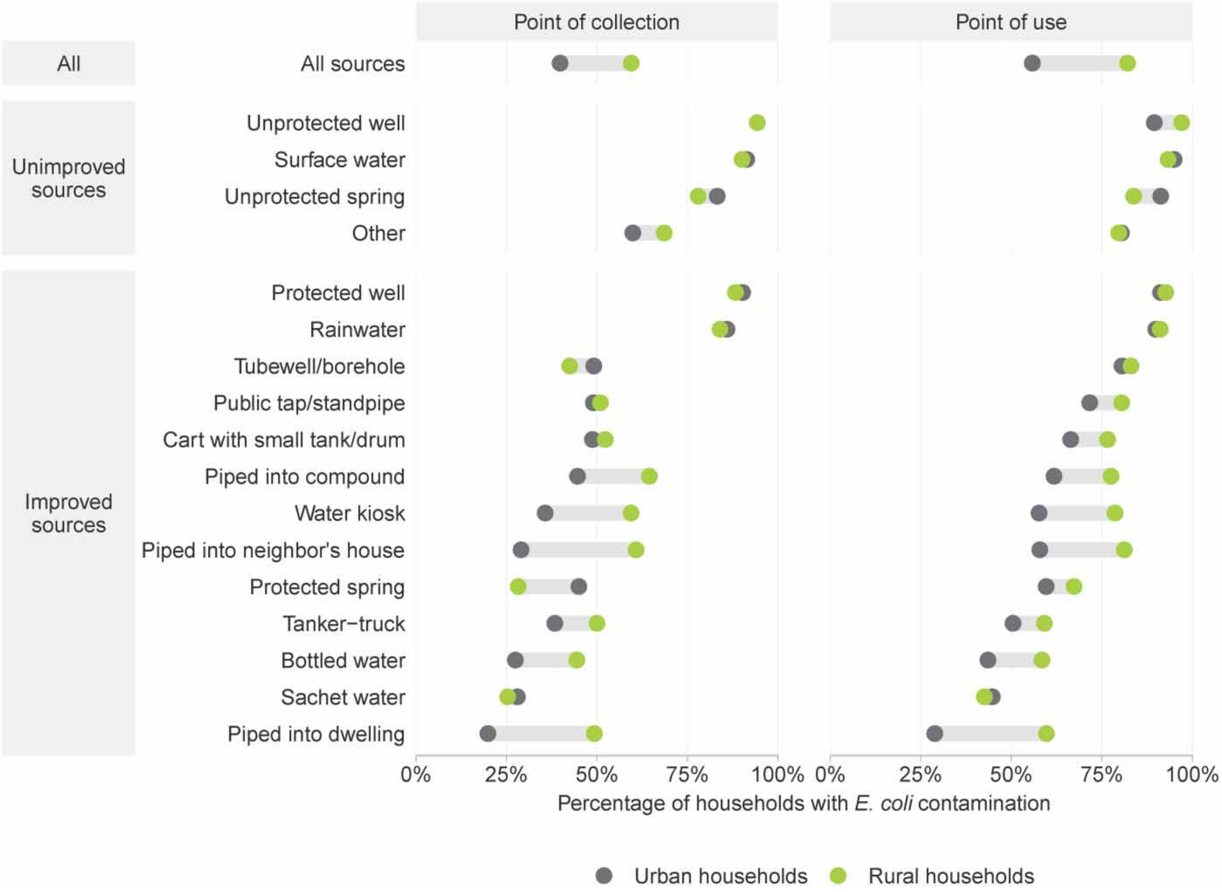

Figure 5 | Comparison in the percentage of households with *E. coli* contamination between urban and rural households. Water sources are ordered according to the pooled percentage of households with *E. coli* contamination at the point of use. In case estimates could not be calculated ($n < 25$) for either the urban or rural setting, and results were not shown for that source.

DISCUSSION

Considering the global target of universal and equitable access to safely managed water by 2030, our results show that *E. coli* contamination was widespread and unacceptably high in almost all water sources, settings, and countries in our sample (predominantly LMICs), with substantial inequalities between and within countries. Some water sources currently classified as improved were as likely to be contaminated as unimproved sources. Furthermore, the capacity of a water source to provide water that did not have *E. coli* contamination was highly context-dependent. Contamination for piped water, in particular, was markedly higher in rural than urban areas. In some countries where contamination was widespread (e.g., Chad, Sierra Leone, and Kiribati), almost all sources had high a likelihood of contamination.

Our results highlight the importance of water quality measurement and how vital it is to measure the last step in providing drinking water. Monitoring activities that use source-centred indicators are strongly reliant on the current classification of improved and unimproved water sources. They also ignore any deterioration in quality that can happen between the source and the point of use. From a public health perspective, monitoring of water quality at the point where people drink the water, instead of at the source, is a better indicator of the health risks associated with water contamination and is an indispensable step towards achieving SDG 6. For example, in Chad, 50.9% of the population had access to basic water services in 2019, but only 1.4% had access to a safely managed water service (INSEED and UNICEF 2020). This means that half of the population had access to an improved source with a collection time of 30 min or less, but almost no one had access to drinking water that was free from faecal contamination (*E. coli*), accessible on the premises and available when needed. The main driver of this difference was water contamination (INSEED and UNICEF 2020), which if not measured leads to a serious overestimation of progress. We urge other nationally representative surveys to include water quality measurement as a standard module for LMICs. This is especially important for the Demographic and Health Surveys and the Living Standards Measurement Study, given that the MICS module has already been adapted to be used for both (JMP 2023).

Classification of water sources

Even though relying on water source for monitoring has serious drawbacks, more than half of the world's population still lack water quality data and depend on the basic water service indicator (WHO and UNICEF 2021). Our findings showed that there were important differences in contamination between improved and unimproved sources. Nevertheless, they also highlight that some sources' classifications need further scrutiny, in particular protected wells and rainwater.

Protected and unprotected wells had a very similar likelihood of contamination, which suggests that protected wells (as currently defined and measured) fail to provide enough protection against *E. coli* contamination. Wells, even when protected against surface water runoff and harmful materials that can fall inside, are still susceptible to contamination via groundwater flow. This can happen when wells are located close to sanitation infrastructures such as pit latrines – where contaminants can travel through the ground – especially in areas with high population density and a large number of latrines (Kiptum & Ndambuki 2012).

The high likelihood of contamination for rainwater emphasises the challenge of properly handled rainwater harvesting systems, especially in lower income settings. In many areas of the Caribbean, southeast Asia, and the Indian and Pacific Oceans – especially in low-lying islands – many communities intensively rely on rainwater for drinking and cooking (Bailey *et al.* 2018). Although initially free from contamination, rainwater becomes progressively more contaminated as it stays in contact with the atmosphere and the structures used for water collection and storage. Faecal contamination is quite common and dependent on the design, structure, materials, maintenance, and the weather (WHO 2022). Based on our results, a single rainwater category classified as improved may not be the best approach for household surveys. It might be beneficial to use separate 'protected rainwater' and 'unprotected rainwater' categories – similar to springs – taking into account the presence of automatic diverters, detachable downpipes, wire meshes, inlet filters, appropriate roof material, and/or storage tank covers as indicators of harvesting systems that limit water contamination.

Since the Millennium Development Goals era, the JMP has updated both its definition of improvement and what sources are classified as improved (Bartram *et al.* 2014). An improved source used to be defined as a source that 'by the nature of its construction and design adequately protects the source from outside contamination, in particular by faecal matter' (WHO 2017). The current definition does not mention contamination, instead focusing on the 'potential to deliver safe water' (WHO, UNICEF and World Bank 2022). This definition is broader, but also more ambiguous, since there is no universally recognised definition of safe water (Dinka 2018). On the one hand, it is more aligned with the WHO's delineation that an adequate drinking water supply should provide not only quality but also accessibility, quantity, continuity, and affordability (WHO 2022). On the other hand, its ambiguity makes the interpretation of global indicators that rely on it less clear. It raises important questions on what specific criteria to include when defining a source as improved and how they should be weighted when they contradict each other. In order to improve public accountability, the JMP could include a clearer definition of 'safe water' and the criteria being used for source classification in their website and their technical reports. Further studies are also necessary to evaluate water sources in terms of their potential to simultaneously provide accessibility, quantity, quality, continuity, and affordability of water supply. Nevertheless, if contamination is the main criteria, there is overwhelming evidence that neither protected wells nor rainwater should remain as improved sources.

Water quality deterioration between the source and the point of use

For many water sources, the prevalence of contamination at the point of use was higher than the prevalence at the point of collection. This was expected, given that contamination can occur during the many stages of water handling and storage that happen between the point of collection and the point of use, especially in LMICs with a deficient sanitation and hygiene infrastructure (WHO and UNICEF 2021). A study conducted in Malawi in 2019 investigated *E. coli* contamination in water in four different stages of water collection: at the water source, at the collection container, at the storage container, and at the cup of drinking water. That study found that the level of contamination would increase in every stage, but the critical steps were filling the collection container and during water storage (Cassivi *et al.* 2021). Most MICS surveys with the water quality module investigate water storage via only one question that classifies households on whether water is collected directly from the source or from a covered or uncovered container. There is an opportunity for more categories and questions to be included that can better differentiate and investigate the water collection, transportation, handling, and storage stages. They also need to be investigated simultaneously with sanitation and hygiene infrastructure to provide evidence for comprehensive WASH interventions that can effectively reduce contamination and achieve major impacts in public health (Cumming *et al.* 2019).

Contamination in urban and rural areas

For many improved sources, the prevalence of contamination was higher in rural areas when compared to urban areas. That difference was particularly high for piped water sources. Therefore, indicators that rely on the classification of improved/unimproved sources or that are based on access to piped water – without measuring water contamination – tend to overestimate the quality of water infrastructure in the rural context. This needs to be taken into account in the interpretation of equity analysis of those indicators. For example, in a region where the urban and rural areas have the same prevalence of access to piped water, the situation is likely to be worst in the rural areas because faecal contamination is likely higher. This is in line with the multiple structural challenges faced by rural populations. Lower population density, larger distances between households, and the lack of political will and resources make the installation and maintenance of WASH infrastructure less feasible, resulting in higher risk of water contamination (Abrams *et al.* 2021; Bain *et al.* 2021; Apatinga *et al*. 2022). In 2020, while 64% of the global urban population had access to sewer connections, only 15% of the rural population had similar access (WHO and UNICEF 2021). Furthermore, farming and animal husbandry can create competition for water resources and result in contamination with agrichemicals and animal waste. Likewise, climate seasonality and extreme weather events exacerbated by climate change can reduce water availability and increase contamination, especially when coupled with improper sanitation (Abrams *et al.* 2021; Apatinga *et al*. 2022).

Limitations and strengths

There are important limitations to our research. First, we only investigated *E. coli* contamination and did not include other important chemical contaminants. This was due to lack of data availability and also the fact that there are region-specific contaminants that are prioritised in different surveys and their inclusion would result in the lack of comparability between national estimates. Second, we focused on the prevalence of any *E. coli* contamination but did not evaluate the level of contamination in the samples (based on the number of colonies found after incubation). This choice was made to align our investigation with SDG 6, which targets the complete absence of water contamination.

Third, the similar likelihood of contamination between protected and unprotected wells might be influenced by misclassification error, i.e., the interviewees might be misclassifying protected wells as unprotected wells and *vice versa*. In the field, there is not always a clear cut difference between these types of wells. Different materials, shapes, and structures can be used, and they offer different levels of protection and are found in different states of repair. Even though interviewees are assisted by the interviewers, errors are still possible and would introduce bias towards similar likelihoods of contamination. A 2019 study in Kenya showed strong inter-observer agreement in classifying water sources, but some sources were more likely to be misclassified, including protected and unprotected wells (Okotto-Okotto *et al.* 2020). Nevertheless, it is unlikely that misclassification alone would be responsible for this pattern, given the fact that the same was not observed for protected and unprotected springs. Protected springs were significantly less likely to be contaminated than unprotected ones. If misclassification is happening for springs, we would expect the difference to be even larger, therefore providing compelling evidence that spring protection is strongly associated with lower risk of contamination. The same cannot be said for wells.

Fourth, even though cross-contamination identified by blank tests was low in most surveys, it was higher than 5% in three countries. A previous study with a similar sample found that excluding countries and clusters with high contamination in the blank tests had only a negligible impact in the final results (Bain *et al.* 2021). Fifth, although we had a large number of households and countries, our sample is not representative of the whole world or of all LMICs. As more surveys become available, these analyses need to be updated to increase the external validity of their results. Despite these limitations, our analyses provide robust evidence of water contamination according to water sources based on a highly comparable multicountry sample stratified by urban and rural settings.

CONCLUSIONS

Our results have shown the pervasiveness of *E. coli* contamination in drinking water sources. Immediate change is necessary to guarantee universal access to water that is free from contamination and from where water is first sourced to the glass where people drink it. There are many opportunities to improve global monitoring of SDG 6 and it would be beneficial (1) to increase the number of nationally representative surveys that include a water quality module, in particular, in Demographic and Health Surveys (DHS) and Living Standards Measurement Study (LSMS) surveys; (2) to present clearer definition, criteria, and evidence used for classifying water sources as improved or unimproved; (3) to further investigate water sources according to those multiple criteria and update the classification of wells and rainwater; and (4) to expand the current investigation of water collection, transportation, handling, and storage in the water quality module of MICS. Monitoring water contamination is essential if we want to achieve SDG 6, especially in rural settings where other simpler indicators might be overestimating the level of the development of water supply systems.

ETHICS STATEMENT

The organisations who administered the surveys were responsible for ethical clearance according to the norms of each country.

FUNDING

This paper was made possible with funds from the Bill & Melinda Gates Foundation (Grant Number: OPP1148933), the Wellcome Trust (Grant Number: 101815/Z/13/Z), the Associação Brasileira de Saúde Coletiva, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Finance code: 001). M.A.B.'s time is supported by an Australian Research Council Discovery Early Career Researcher Award (DE200100264) and a Dame Kate Campbell Fellowship.

DATA AVAILABILITY STATEMENT

All the analyses were carried out using publicly available datasets that can be obtained directly from the MICS website (mics.unicef.org). Datasets are continuously sourced and updated by the International Center for Equity in Health (equidade.org) as they are released. We used the last available versions on 8 February 2023.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abrams, A. L., Carden, K., Teta, C. & Wågsæther, K. 2021 Water, sanitation, and hygiene vulnerability among rural areas and small towns in South Africa: Exploring the role of climate change, marginalization, and inequality. *Water* **13**, 20. https://doi.org/10.3390/w13202810.

Apatinga, G. A., Schuster-Wallace, C. J. & Dickson-Anderson, S. E. 2022 A conceptual framework for gender and climate mainstreaming to mitigate water inaccessibility in rural sub-Saharan Africa. *Wiley Interdisciplinary Reviews – Water*. https://doi.org/10.1002/wat2.1591.

Bailey, R. T., Beikmann, A., Kottermair, M., Taboroši, D. & Jenson, J. W. 2018 Sustainability of rainwater catchment systems for small island communities. *Journal of Hydrology* **557**, 137–146. https://doi.org/10.1016/j.jhydrol.2017.12.016.

Bain, R., Cronk, R., Wright, J., Yang, H., Slaymaker, T. & Bartram, J. 2014 Fecal contamination of drinking-water in low- and middle-income countries: A systematic review and meta-analysis. *PLoS Medicine* **11** (5), e1001644. https://doi.org/10.1371/journal.pmed.1001644.

Bain, R., Johnston, R., Khan, S., Hancioglu, A. & Slaymaker, T. 2021 Monitoring drinking water quality in nationally representative household surveys in low- and middle-income countries: Cross-sectional analysis of 27 multiple indicator cluster surveys 2014–2020. *Environmental Health Perspectives* **129** (9), 97010. https://doi.org/10.1289/EHP8459.

Bartram, J., Brocklehurst, C., Fisher, M. B., Luyendijk, R., Hossain, R., Wardlaw, T. & Gordon, B. 2014 Global monitoring of water supply and sanitation: History, methods and future challenges. *International Journal of Environmental Research and Public Health* **11** (8), 8137–8165. https://doi.org/10.3390/ijerph110808137.

Cassivi, A., Tilley, E., Waygood, E. O. D. & Dorea, C. 2021 Household practices in accessing drinking water and post collection contamination: A seasonal cohort study in Malawi. *Water Research* **189**, 116607. https://doi.org/10.1016/j.watres.2020.116607.

Cumming, O., Arnold, B. F., Ban, R., Clasen, T., Esteves Mills, J., Freeman, M. C., Gordon, B., Guiteras, R., Howard, G., Hunter, P. R., Johnston, R. B., Pickering, A. J., Prendergast, A. J., Prüss-Ustün, A., Rosenboom, J. W., Spears, D., Sundberg, S., Wolf, J., Null, C., Luby, S. P., Humphrey, J. H. & Colford Jr, J. M. 2019 The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: A consensus statement. *BMC Medicine* **17** (1). https://doi.org/10.1186/s12916-019-1410-x.

Dinka, M. O., 2018 Safe drinking water: Concepts, benefits, principles and standards. In: (Glavan, M., ed.). IntechOpen, Rijeka, Ch. 10. https://doi.org/10.5772/intechopen.71352.

INSEED and UNICEF 2020 *MICS6-Tchad, 2019, Rapport final*. N'Djamena.

JMP 2023 *Water Quality Monitoring*. Available at: https://washdata.org/monitoring/drinking-water/water-quality-monitoring (accessed 25 August 2023).

Khan, S. & Hancioglu, A. 2019 Multiple indicator cluster surveys: delivering robust data on children and women across the globe. *Studies in Family Planning* **50** (3), 279–286. https://doi.org/10.1111/sifp.12103.

Kiptum, K. & Ndambuki, J. 2012 Well water contamination by pit latrines: A case study of Langas. *International Journal of Water Resources and Environmental Engineering* **4** (2), 35–43. https://doi.org/10.5897/IJWREE11.084.

Lumley, T. 2004 Analysis of complex survey samples. *Journal of Statistical Software* **9** (8), 1–19. https://doi.org/10.18637/jss.v009.i08.

Okotto-Okotto, J., Wanza, P., Kwoba, E., Yu, W., Dzodzomenyo, M., Thumbi, S. M., da Silva, D. G. & Wright, J. A. 2020 An assessment of inter-observer agreement in water source classification and sanitary risk observations. *Exposure and Health* **12** (4), 809–822. https://doi.org/10.1007/s12403-019-00339-3.

UN General Assembly 2010 *The Human Right to Water and Sanitation: Resolution/Adopted by the General Assembly*. UN, New York.

UNICEF 2016 *MICS6 Tools | MICS Manual for Water Quality Testing*. Available at: https://mics.unicef.org/tools?round=mics6#data-collection (accessed 22 May 2023).

UNICEF 2020 *MICS6 Tools | Household Questionnaire*. Available at: https://mics.unicef.org/tools?round=mics6 (accessed 22 May 2023).

UNICEF 2021a *MICS6 Tools | MICS Instructions for Interviewers*. Available at: https://mics.unicef.org/tools?round=mics6#data-collection (accessed 25 May 2023).

UNICEF 2021b *MICS6 Tools | Survey Findings Report & Snapshot Guidelines*. Available at: https://mics.unicef.org/tools?round=mics6#reporting (accessed 24 May 2023).

UNICEF 2023 *MICS Surveys*. Available at: https://mics.unicef.org/surveys (accessed 8 February 2023).

UNICEF and WHO 2018 *Core Questions on Drinking Water, Sanitation and Hygiene for Household Surveys: 2018 Update*. New York, World Health Organization and UNICEF.

UNICEF and WHO 2020 *Integrating* Water Quality Testing *into* Household Surveys*: Thematic* R*eport on* Drinking Water. New York, World Health Organization and UNICEF.

United Nations 2017 *Principles and Recommendations for Population and Housing Censuses, Revision 3*. United Nations, New York.

United Nations 2023a *Goal 6: Targets and Indicators*. Available at: https://sdgs.un.org/goals/goal6 (accessed 5 September 2023).

United Nations 2023b *SDG Indicators: Regional Groupings Used in Report and Statistical Annex*. Available at: https://unstats.un.org/sdgs/indicators/regional-groups/ (accessed 25 May 2023).

WHO 2017 *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*. Geneva, World Health Organization.

WHO 2022 *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First and Second Addenda*. Geneva, World Health Organization.

WHO and UNICEF 2021 *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years into the SDGs*. Geneva, World Health Organization and UNICEF.

WHO, UNICEF and World Bank 2022 *State of the World's Drinking Water: An Urgent Call to Action to Accelerate Progress on Ensuring Safe Drinking Water for All*. Geneva, World Health Organization, UNICEF and World Bank.

World Bank 2022 *Population, Total*. Available at: https://data.workldbank.org/indicator/SP.POP.TOTL (accessed 8 February 2023).

First received 19 June 2023; accepted in revised form 28 October 2023. Available online 9 November 2023