Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Estimating the prevalence of water insecurity experiences across 194 countries

Indira Bose

indira.bose@lshtm.ac.uk

London School of Hygiene and Tropical Medicine (LSHTM) https://orcid.org/0000-0002-3045-8851

Joshua D. Miller

University at Buffalo, State University of New York

Esther E Greenwood

Swiss Federal Institute of Aquatic Science and Technology

Pierre Masselot

London School of Hygiene & Tropical Medicine

Karin Gallandat

Swiss Federal Institute of Aquatic Science and Technology

Richard Johnston

World Health Organization

Suneetha Kadiyala

London School of Hygiene & Tropical Medicine

Edward A Frongillo

University of South Carolina

Sera L Young

Northwestern University

Research Article

Keywords: Water Security, Sustainable Development Goals, Global Health, Sustainability, Water, Measurement, Disparities.

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11669560/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Estimating the prevalence of water insecurity experiences across 194 countries

Authors: Indira Bose[1,2*], Joshua D. Miller[3,4], Esther E. Greenwood[5], Pierre Masselot[6], Karin Gallandat[5], Richard Johnston[7], Suneetha Kadiyala[1], Edward A. Frongillo[8], Sera Young[9,10,11*]

Affiliations

1. Department of Population Health, London School of Hygiene & Tropical Medicine, U.K.
2. The Centre on Climate Change and Planetary Health, London School of Hygiene & Tropical Medicine, U.K.
3. Department of Epidemiology and Environmental Health, University at Buffalo, USA.
4. Center for Climate Change and Health Equity, University at Buffalo, USA.
5. Department Sanitation, Water and Solid Waste for Development (Sandec), Swiss Federal Institute of Aquatic Science and Technology (Eawag), Duebendorf, Switzerland
6. Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK
7. Department of Environment, Climate Change, One Health and Migration, World Health Organization, Geneva, Switzerland.
8. Department of Health Promotion, Education, and Behavior, University of South Carolina, USA.
9. Department of Anthropology, Northwestern University, USA.
10. Institute for Policy Research, Northwestern University, USA.
11. Center for Water, Northwestern University, USA.
12. United Nations University-Institute for Water, Environment & Health, Canada.

*Corresponding authors:

Dr Indira Bose, Email: Indira.bose@lshtm.ac.uk;

Professor Sera Young, Email: sera.young@northwestern.edu

Abstract/Summary

Reliable access to sufficient, safe water is a human right and fundamental to health and well-being. Global monitoring has focused on water resources and household drinking-water service coverage, but not whether people can reliably meet their water needs, that is, whether people are water secure. Consequently, the global burden of individuals' experiences with water insecurity remains poorly characterised. We use a hierarchical machine learning approach to provide the first estimates of the global prevalence of individual water insecurity across 194 countries, drawing upon nationally representative data on water insecurity experiences covering 76 countries (79% of the global population ≥15 years). We estimate that approximately 909 million youth and adults globally (15%) experience in 2025 moderate-to-high water insecurity. Prevalence is highest in Africa (36%), although the largest absolute burden is in Asia. Country estimates were frequently discordant with traditional water-supply indicators, such as water stress and coverage of drinking-water services, demonstrating the value of incorporating people's experiences into global assessments of water security.

1. Introduction

Reliable access to water of adequate quality and quantity is a fundamental human right [1]. Disruptions in water quality or availability can trigger cascading effects across food systems, amplify infectious disease risks, and ultimately undermine population health and well-being [2-6]. These challenges are further compounded by ongoing climate hazards, infrastructural failures, governance constraints, as well as increasing demand driven by rapid urbanisation, population growth, and industrial expansion [7-10]. Yet understanding the scale of their consequences for people requires knowing not only whether freshwater resources and water infrastructure are available, but whether individuals can reliably access sufficient, safe water to meet their daily needs.

Water insecurity is defined as the inability of people to reliably access safe and sufficient water to meet basic daily needs [11]. Current global monitoring efforts focus primarily on supply-side metrics, such as per capita freshwater availability or household-use of safely managed drinking-water services [12, 13]. These indicators are currently used to track progress towards Sustainable Development Goal (SDG) 6: Ensure availability and sustainable management of water and sanitation for all [14]. Water stress, measured through comparing freshwater withdrawal to the total supply, provides important insights into the physical availability of water at the national and subnational level [15], but physical availability does not mean that all individuals in an area can reliably access it. Safely managed drinking-water services (SMDWS) incorporates important domains of accessibility, availability and quality of water at the household level, including whether water is available when needed [16]. These measures, however, do not capture the ways in which water problems are experienced or affect people's daily lives, or if the right to water is being met. Moreover, household-level measures cannot capture whether water insecurity varies between individuals within the same household based on gender and other socio-demographic factors [17-19].

Given that water insecurity is linked with numerous adverse physical and mental health outcomes [2, 20], there is a critical need for metrics that capture people's experiences of water insecurity directly and holistically. The Individual Water Insecurity Experiences (IWISE) Scale was developed to meet this need, providing a valid, cross-country equivalent measure for quantifying individual-level experiences with water insecurity [21, 22]. The IWISE Scale offers important nuance by identifying population subgroups whose ability to reliably access safe and sufficient water is most precarious [23]. Despite increasing use of the tool, the global prevalence of and regional variation in water insecurity remain unknown. Such global and regional estimates can provide a critical benchmark against which we can track progress towards achieving water security for all. Estimating these prevalences can also identify the countries and regions that are disproportionately affected and inform policies and resource allocation decisions.

The availability of nationally representative IWISE data collected via the Gallup World Polls (GWP) between 2020 and 2025, together with advances in predictive modelling using machine learning, provide an opportunity to estimate water insecurity experiences globally. In this study, we estimate the prevalence of individual water insecurity across 194 countries by combining these observed nationally-representative IWISE data (from 76 countries representing 79.4% of

the global population) with machine learning predictions to fill data gaps for countries where IWISE data is lacking (118 countries representing 20.6%). We further examine how individuals' experiences with water insecurity correspond with established global indicators of water stress and drinking-water services. Together, these analyses provide the first global assessment of water insecurity.

To generate these global and regional estimates of moderate-to-high water insecurity (referred to as "water insecurity" throughout this study), we used a hierarchical machine learning method to predict estimates for the unobserved countries. Multiple machine learning and traditional statistical methods were tested and validated using the observed data to identify the models that had the best predictive performance. This study therefore also demonstrates a novel application of machine learning methods that utilises a mix of individual- and national-level data (including a range of economic, environmental and sociodemographic indicators *Tables S1-4*) to estimate global, regional, and national prevalences of health determinants or conditions.

2. Results

2.1 Global and regional water insecurity estimates

The observed analytic sample comprised 89,740 individuals from 76 countries, representing 79.4% of the global population aged ≥15 years. Forty-nine percent were women, and the median age was 36 years old (IQR 25–50 years); see *Table S1 for full descriptive statistics*. The observed countries included 39 from Africa, 16 from Asia, 16 from the Americas, 4 from Europe and 1 from Oceania, spanning all World Bank FY27 country income groups: 15 low-income, 30 lower-middle-income, 21 upper-middle-income, and 10 high-income.

The prevalences of water insecurity for 194 countries were estimated by combining observed prevalences with predictions for 118 countries (**Figure 1**). Across these 194 countries, an estimated 14.7% of people aged 15 years and older experienced water insecurity (95% uncertainty interval (UI): 14.0-15.5%), corresponding to approximately 909 million youth and adults (95% UI: 862 - 960 million).

Across country income groups, prevalence was highest in low-income countries (40.7%, 95% UI: 36.4%- 44.4% **Figure 2B**), whereas the largest number of people experiencing water insecurity lived in lower-middle-income countries (422 million, **Figure 2B**).

Regionally, the estimated prevalence of water insecurity was highest in Africa (36.2%, 95% UI: 34.2-37.8%), followed by the Americas (16.9%, 95% UI: 13.9%- 19.9% (**Figures 1** and **2**; *Figure S1*). The largest number of people experiencing water insecurity resided in Asia, approximately 397 million (95% UI: 365-436 million; 10.6%, 95% UI: 9.8%- 11.7%) (**Figure 2A**). When broken down by sub-region (*Figure S2-3*) the largest numbers of people experiencing water insecurity were based in Sub-Saharan Africa (288 million, 95% UI: 274-300 million; 38.0%, 95% UI: 36.1%-39.5%), followed by South Asia (252 million, 95% UI: 292-280 million; 16.4%, 95% UI: 14.9%-18.2%). Although Europe had one of the lowest prevalences of water insecurity (4.2%, 95% UI:

2.9-6.0%), an estimated 27 million people in the region were estimated to experience water insecurity.

2.2 Comparison with key global water monitoring indicators

We compared country-level prevalence of water insecurity experienced by individuals with three established global indicators capturing different dimensions of water from a supply perspective: 1) water stress, defined as freshwater withdrawal as a proportion of available freshwater resources (aligned to SDG monitoring indicator 6.4.2) ; 2) proportion of population lacking at least basic drinking-water services, defined as drinking-water from an improved source, provided the collection time is no more than 30 minutes for a round trip; and 3) proportion of population lacking safely managed drinking-water services (SMDWS), defined as drinking water from an improved source that is accessible on premises, available when needed and free from faecal and priority chemical contamination (SDG monitoring indicator 6.1.1). We conducted bivariate mapping of these indicators with water insecurity prevalence (**Figure 3**) to understand the geographic overlap. We also calculated correlations between each indicator and water insecurity prevalence, (**Figure 4**) and partial correlations accounting for gross national income per capita (GNI) (*Figure S4*). We also calculated these correlations among countries with observed water insecurity data only ("observed only"), as water stress and household-use of basic drink-water services were predictors in the models, correlations including predicted countries ("overall") may be higher for these two indicators (*Figure S5-6*).

Water insecurity was weakly correlated with water stress (overall r=-0.15, observed only r=-0.21, *Figure S5*). There was substantial discordance between countries with a high prevalence of water insecurity and those classified as having high levels of water stress. For example, countries in Sub-Saharan Africa such as Chad and Cameroon had some of the highest prevalence estimates of water insecurity (using observed data: 66.3% and 63.9% respectively) but had relatively low physical water stress (4.3% and 1.6% respectively). Conversely, in the US and Europe, some countries had moderately high water stress but low prevalence of water insecurity. For example, in the US and the UK, water insecurity prevalence was relatively low (using observed data: 3.7% and 0.5% respectively), yet water stress was 28.2% and 14.4%, respectively.

There was a stronger correlation between water insecurity with both lack of coverage of basic and safely managed drinking-water services (basic r=0.70, observed only r=0.60; SMDWS r=0.65, observed only r=0.60). Although there was a similar overall trend observed with these service level indicators, there was still substantial discordance between these indicators in numerous countries. In Latin America, in particular, there were many countries with relatively high water-insecurity prevalence despite high coverage of basic drinking-water services (i.e. low proportion without access to these services), whereas the inverse was the case in Southeast Asia (**Figure 3**). Although the overall correlation was lower with SMDWS, for most countries where SMDWS data were available, those identified as having a high-water insecurity prevalence were found to have low access to SMDWS. In Sub-Saharan Africa there was high overlap in the countries identified as having a high prevalence of both water insecurity and lack of basic or safely managed drinking-water services.

To better unpack the relationship with SMDWS, we additionally calculated the correlation between water insecurity and each of the three components of safely managed drinking-water separately: availability when needed, free from contamination and water source available on premises (**Figure 4** C1-C3, and *Figure S4-6* observed only and partial correlations GNI adjusted). Water insecurity was most closely associated with water not available when needed (r=0.62, observed only r=0.66) and water not on premises (r=0.70, observed only r=0.58). The correlation with water not free from contamination was weaker (r=0.57, observed only r=0.53). There was strong overlap in the countries found to have high prevalence of water not available when needed and those with high water-insecurity prevalence. Nevertheless, there were some exceptions, such as Brazil, which had a relatively higher water-insecurity prevalence (16.1%, observed data) but low proportion of people without water available when needed (1.1%, *Figure S7*). Conversely, in Indonesia, water-insecurity prevalence was estimated to be 5.3% using observed data, but 13.2% of the population did not have water available when needed.

In the partial correlations, accounting for GNI, the associations between water insecurity and all other water indicators were generally attenuated (*Figures S4 & 6*). Although weak positive correlations remained with lack of access to basic drinking water (r=0.35, observed only r=0.33), water not available needed (r=0.35, observed only r=0.53), and water not on premises (r=0.34, observed only r=0.27).

2.3 Model validation

For the 118 countries without observed IWISE data, we predicted water-insecurity prevalence using national- and individual-level data where available ("individual-level model", n=85 countries, representing 20% of the global population aged ≥15 years **Figure 5D**), and national-level data alone ("country-level model", n=33 countries, representing 0.6% the global population aged ≥15 years **Figure 5D**).

We tested multiple machine learning models and traditional statistical methods, with Elastic Net showing the best predictive performance for both individual- and country-level models (*see Methods and Figure S8-13 and Table S5-8 for more details*). Performance was stronger for the individual-level model than the country-level model (R^2 = 0.69 vs 0.56 and MAE =0.07 vs 0.09), supporting the use of individual-level predictions wherever possible. Overall, the individual-level Elastic Net model demonstrated good calibration in the Leave-Countries-Out Cross-Validation (LCO CV) (**Figure 5**), with minimal evidence of systematic bias across deciles of predicted prevalence or by region (**Figure 5,** *Figure S11*) and income group (*Figure S12*). The country-level model demonstrated weaker calibration, although there was similarly minimal evidence of systematic bias across the groups examined (*Figure S11-13*).

Countries with greater prediction uncertainty were predominantly fragile states and Pacific Island countries, such as Haiti, Sudan, Papua New Guinea, Solomon Islands, Kiribati, Micronesia, Marshall Islands, Timor-Leste and Tuvalu (*Table S11-12*). Predictions for these countries were flagged as potentially being less robust because their predictor data were more limited, older, and often outside the range of the training data, or prediction estimates differed substantially between models.

2.4 Key Predictors

The final individual- and country-level models included 11 national-level variables (e.g. GNI) and region as candidate predictors. The individual-level model included an additional 13 individual-level variables (e.g. age), including three missing-data flags, as candidate predictors. These variables were selected as they could be plausibly associated with water insecurity based on prior research and were available for most countries (*see 4. Methods*). To aid interpretation, we group these variables as: water (e.g. access to basic drinking water, water stress, satisfaction with water quality at the individual-level); economic (e.g. GNI, income quintile); sociodemographic (e.g. age, gender); geographic (e.g. region, urbanicity); climate & environment (e.g. climate vulnerability index, precipitation); health (e.g. life expectancy, fertility rate); and governance (public sector corruption index). Two of the section 2.2 comparison indicators, access to basic drinking water and water stress, were included as predictors. SMDWS and its components, including availability when needed, were excluded because of missing data.

Satisfaction with water quality was the strongest predictor in the final individual-level Elastic Net model (**Figure 6**, see *Table S9 for full list of predictors selected*). As this variable is a subjective measure of a related construct, we refitted the models without this predictor. Performance declined only modestly (R^2 0.69 to 0.62; MAE 0.070 to 0.080), indicating that the model retained substantial predicted performance without this indicator (*Figure S16-17*). Region was also an important predictor in both the individual- and country-level Elastic Net models, with residence in Latin America associated with higher predicted risk of water insecurity. In these models, higher coverage of basic drinking-water services was associated with lower predicted risk of water insecurity, and higher water stress with higher predicted risk of water insecurity. Higher ND-GAIN scores (lower climate vulnerability) and higher corruption index scores (lower perceived corruption) were also associated with lower predicted risk of water insecurity. Feature importance in XGBoost, used as a comparator, was broadly similar. Region carried more importance in Elastic Net, and individual characteristics such as age and household size carried more importance in XGBoost (*Figure S15; country-level comparisons in Figure S18-19*).

Overall, the most important predictors spanned WASH, regional, governance and climate-vulnerability indicators (see *Tables S9-10 for full list of predictors in the final individual and country-level model*). These are predictors used in machine learning prediction models, so should not be interpreted as independent associations or determinants of water insecurity.

3. Discussion

Here we present the first global estimates of water insecurity experiences across 194 countries. An estimated 909 million youth and adults, approximately 14.7% of the global population in this age group, experience in 2025 moderate-to-high water insecurity. This burden is unequally distributed, with the highest prevalence in low-income countries and Africa, while absolute numbers are largest in lower-middle-income countries and Asia.

We generated these estimates by combining nationally representative data from 76 countries with modelled estimates for an additional 118 countries, using a novel hierarchical machine learning method that leverages both individual- and national-level data. The Elastic-Net models used for predictions demonstrated good predictive performance, in particular the individual-level models that used both individual- and national-level data. These observed and predicted estimates provide a new global picture of water insecurity that centres people's experiences meeting their essential water needs, complementing assessments of water availability, access and use, based on supply or service-level indicators.

The substantial discordance between the prevalence of water insecurity experiences and the national level of physical water stress, measured as freshwater withdrawal relative to available resources, highlights the distinction between environmental water-resource constraints and whether those constraints are ultimately experienced by individuals. High or moderately high water stress at national level did not necessarily translate into high prevalence of water insecurity, e.g., in the US, while high water insecurity prevalence was also observed in countries with relatively low water stress nationally, such as in Chad. These findings suggest that the relationship between physical water availability and individuals' experiences with water insecurity is at least partially mediated by the systems through which water is managed, distributed and accessed [24]. Adequate water-resource infrastructure [25] and effective water management systems [26, 27] therefore may help to buffer populations from the challenges of limited freshwater availability and reduce water insecurity, whereas inadequate services and barriers to access may result in people experiencing water insecurity even where freshwater resources are less constrained. These findings clearly suggest that water-resource availability alone may not well reflect people's experiences of water insecurity. Future research should test the mechanisms through which water-resource availability may affect water insecurity.

Water insecurity was more strongly associated with drinking-water service levels than with water stress, but there was still substantial discordance in many countries, particularly after adjusting for GNI. Lack of basic drinking-water services tended to be less prevalent than water insecurity, which is likely because this indicator does not capture challenges such as water supplies being intermittent, insufficient or unaffordable. This divergence was particularly marked in Latin America, where water insecurity prevalence was high despite high basic drinking-water service coverage. Residence in Latin America was also an important predictor of water insecurity in both the Elastic Net and XGBoost models, which included service-level and socioeconomic indicators, suggesting that the available predictors may not fully capture the factors associated with water insecurity in this region. These regional differences could be due to challenges related to intermittent supply, affordability constraints and access inequities within countries in Latin America [28, 29]. These mechanisms warrant investigation in future research.

Conversely, lack of SMDWS tended to be more prevalent than water insecurity, though not consistently. Absence of faecal contamination was the element of SMDWS least associated with water insecurity, possibly because water contamination is typically invisible to users [30]. Water insecurity was slightly more closely associated with availability when needed and accessibility on premises. Access to water on premises does not guarantee that the water supply is sufficient or reliable, and people without water accessible on premises or that is

always available when needed may meet their needs through a variety of coping strategies such as seeking alternative sources [18, 31-33]. In addition, availability when needed is estimated from household surveys in some countries and from utility data on service continuity in others. Neither of these indicators capture how frequently individuals experience water problems, which is captured through measures such as IWISE. There are some caveats to these comparisons. For one, these analyses compare lack of SMDWS with moderate-to-high water insecurity. SMDWS sets a high threshold for this service level, whereas our estimates exclude people experiencing low water insecurity. Including low levels of water insecurity would increase estimated water insecurity prevalence and likely narrow the gap between the two measures in countries where lack of SMDWS is more prevalent. Second, because these comparisons are at country level and based on different data sources, we cannot determine whether those lacking services are the same individuals experiencing water insecurity.

Nevertheless, these findings reinforce that service-based and experiential measures capture different aspects of water problems. More broadly, the discordance between IWISE and other global water monitoring indicators (water stress and drinking-water services) suggests that none of these indicators reliably reflect the full extent of unmet water needs. Measuring experiential water insecurity can therefore complement supply- and service-based monitoring, helping to identify who is affected, how deeply, and where they are located, to track progress and prioritise action.

Water insecurity, as measured by IWISE, captures how water issues affect people's daily lives, including people's ability to meet their water needs for hygiene (e.g. handwashing and bathing), drinking, food preparation and the psycho-social manifestations of these experiences (e.g. worry, anger and shame) [21]. Water insecurity has been associated with multiple adverse physical and mental health outcomes [2, 20, 34], including food insecurity [35], infectious disease [36] and probable depression [37]. Water insecurity experiences have also been identified as an important potential mediator between climate and health [3]. Therefore, in addition to supporting global monitoring, capturing these experiences is critical for better understanding health and well-being, particularly given global stressors such as climate change, conflict and economic volatility that may heighten water insecurity [8, 38, 39].

The methodological techniques herein, combining observed data with predictive estimates from Elastic Net machine learning models, may be useful for subsequent estimates of global public health burdens. The validation and calibration of these machine learning models indicated good predictive performance outperforming the alternative machine learning and traditional statistical methods evaluated in this study. Traditional statistical methods have underpinned the estimation of several global health indicators, but our findings suggest that machine learning methods may offer a useful complement when data are sparse, much as heteroscedastic penalized longitudinal models have been developed to estimate annual prevalence of child malnutrition given that country annual data are sparse [40]. The individual-level model used in this study compared favourably with machine learning models used to estimate other water indicators [41]. This modelling highlights the importance of predictors capturing individual- or household-level characteristics, such as income and age, alongside national-level data to be able to accurately predict health indicators when data are not available.

Although the predictive performance of these models enables us to estimate prevalence for countries for which data are currently absent, making it possible to assess global and regional prevalence without waiting for universal data collection, there are some limitations. Confidence is uneven, particularly for countries relying on country-level prediction alone and some fragile and data-sparse settings, which calls for caution when interpreting these estimates at country level and has implications for both research and policy. For one, data collection of water insecurity experiences should be targeted strategically towards countries where predictions are most uncertain (see *Table S11*) to both improve the global estimates and allow for further validation and fine-tuning of the models. Further, modelled estimates can help identify populations and countries that are potentially overlooked, but these predictions should not be treated as equivalent to directly observed prevalence. They can help to prioritise and complement direct measurement rather than substitute for it.

Despite these limitations, we have confidence in the global models because the global estimates are supported by directly observed data covering 79% of the global population (aged ≥15 years), while the country-level model was used for only 0.6% of the population. Similarly, regional estimates were based predominantly on observed data and data from the individual-level model, with the exception of Oceania (see Figure 2), and therefore uncertainty is highest for this region and future data collection should be prioritised for this region.

There are several avenues for future work that can enhance our understanding of the global picture of water insecurity. First, the estimates presented in the study are of moderate-to-high water insecurity and therefore provide a conservative picture of water insecurity; extending these estimates to also capture those experiencing low water insecurity can give a more holistic view of the breadth of these issues. Also examining how these estimates vary by individual-level characteristics such as age and sex/gender can help us to understand how this burden may vary across different sub-groups.

Second, this study focuses on youth and adults (≥15 years). It is not clear, however, how water insecurity of this age group directly relates to children's experiences, particularly in high-burden countries where children comprise a substantial proportion of the population: 46% and 41% of the population of Chad and Cameroon are under 15 years, respectively. In some instances, children may be buffered from some experiences if their access to water is prioritised [18]; alternatively, in many settings, children are required to fetch water and may experience additional physical and emotional strain related to water insecurity [42-44]. Evidence from food insecurity research indicates that caregivers may underestimate children's experiences [45], underscoring the need for the direct measurement of water insecurity experiences among children to better understand the severity and prevalence, as well as the consequences for their health and well-being.

Third, the estimates presented here provide a picture of the national prevalence of water insecurity, but this may mask substantial subnational inequities and differences between population groups [46, 47]. The inclusion of experiential water measures in routine surveys, such as the World Bank's Living Standards Measurement Study or other national surveys would reveal subnational variation. Also, while the IWISE Scale has been validated in low-, middle- and high- income settings [21, 22, 48], and has also been used and validated in countries beyond

the presented observed data [49-52], countries may seek to further understand how these experiences may manifest in their contexts.

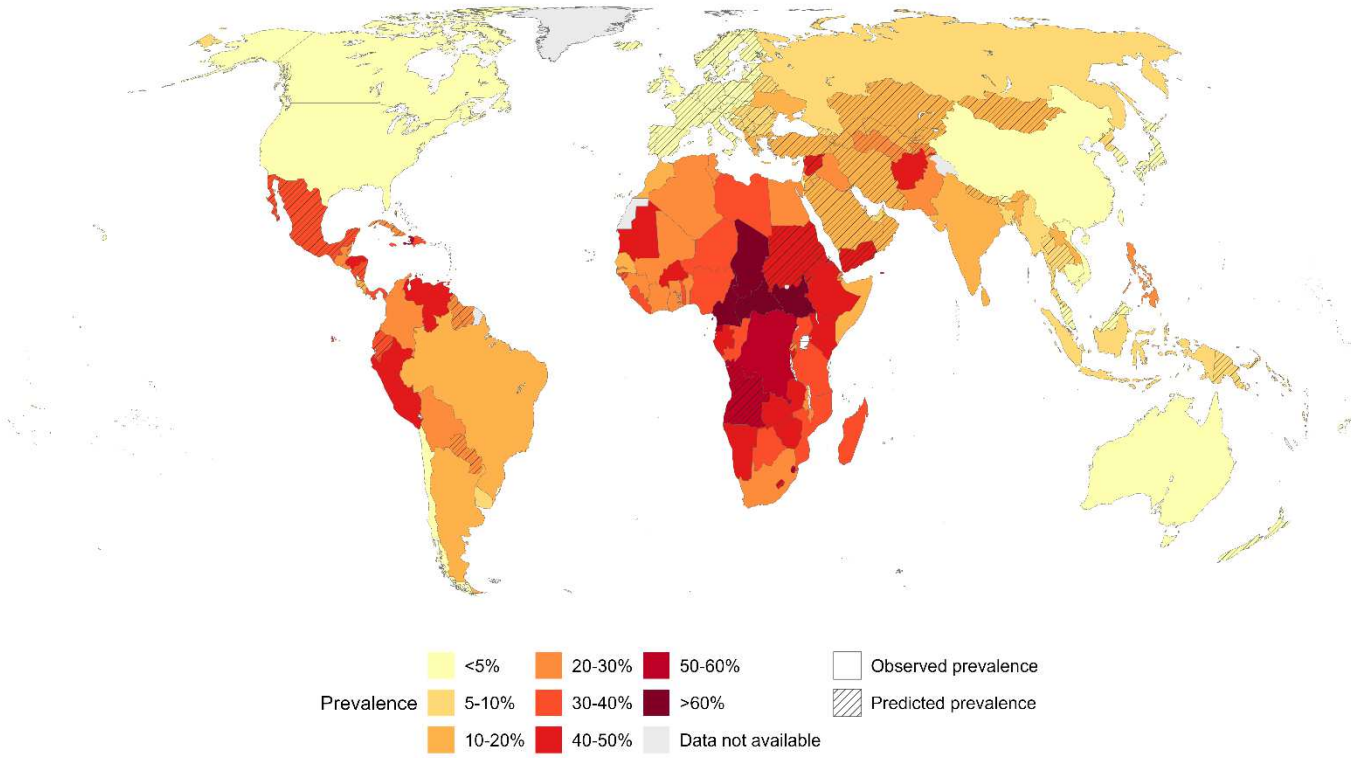
Fourth, we estimated overall water insecurity prevalence, based on a 12-month recall period, using 2025 populations projections to estimate the absolute burden. However, water insecurity is not static and may vary across time. Future research into seasonal and annual trends across heterogenous settings, supported by repeated high frequency measurement, would be another valuable area for future research, especially in the context of climate and economic shocks.

Finally, the key predictors used by these models are not necessarily determinants of water insecurity. Identifying the mechanisms through which water-resource availability, infrastructure, service reliability, socioeconomic and climatic conditions shape people's experiences, and evaluating interventions that address these mechanisms, will be essential to identify effective action to reduce the burden of water insecurity presented in this study.

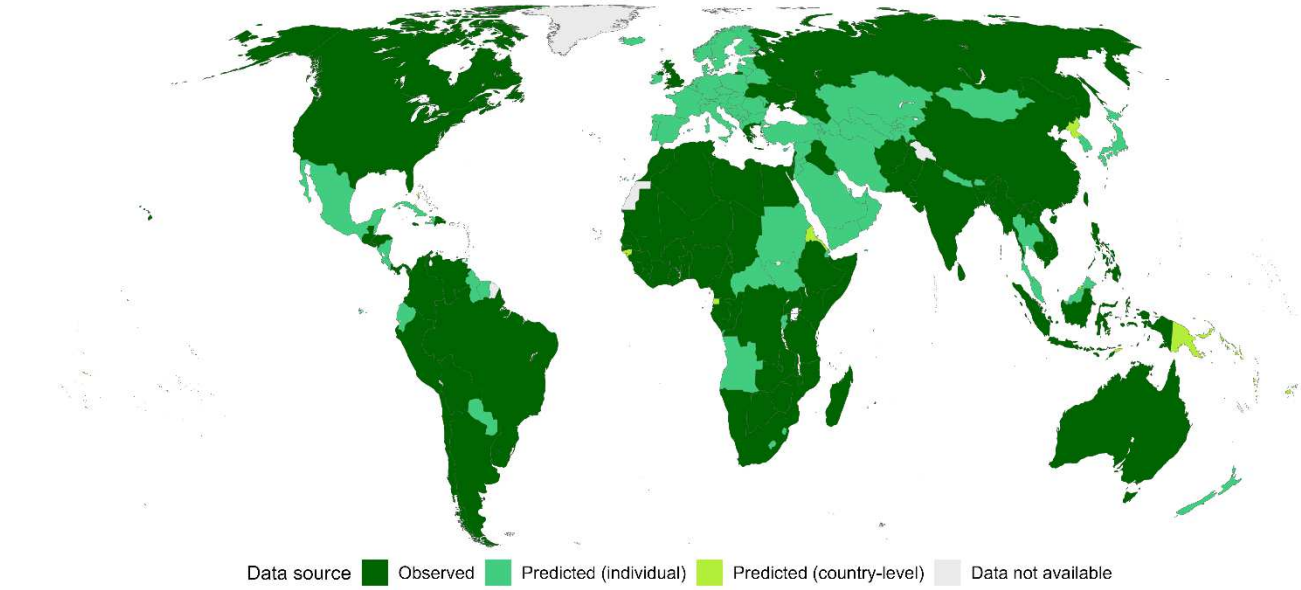
In conclusion, by estimating water insecurity experiences across 194 countries, we provide a new global evidence base on the extent and distribution of people’s difficulties in meeting their water needs. Our findings demonstrate existing indicators measuring freshwater availability and drinking-water service levels do not fully capture the geographical distribution of water insecurity experiences. Incorporating water insecurity experiences into global monitoring alongside existing indicators could provide a more complete picture of where people's essential water needs remain unmet, to inform priorities for data collection, policy and resource allocation more effectively. Importantly, these global estimates also provide a basis for examining how water insecurity experiences relate to health, nutrition and well-being across diverse contexts, while helping identify settings where direct measurement and policy responses are most needed and supporting progress tracking over time.

Figure 1 Map of Estimated Global Prevalence of Water Insecurity* *Panel A: Map of estimated prevalence of youth and adults experiencing moderate to high water insecurity in 194 countries; Panel B data source*

a) Estimated water insecurity prevalence



b) Data source



**The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the authors concerning the legal status of any country, territory, city or area or of its authorities, or concerning, the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which, there may not yet be full agreement.*
Map boundaries: UN Geospatial; Abyei, Kosovo and lakes: Natural Earth. Projection: Equal Earth.

Figure 2 Bar Chart of Prevalence of Water Insecurity by A-Region *(Panel A1: estimated prevalence of youth and adults experiencing moderate to high water insecurity by region and globally; Panel A2: total number of youth and adults experiencing moderate to high water insecurity by region and globally)* **and B-Country Income Group** *(Panel B1: estimated prevalence of youth and adults experiencing moderate to high water insecurity by country income group and globally; Panel B2: total number of youth and adults experiencing moderate to high water insecurity by country income group and globally)*

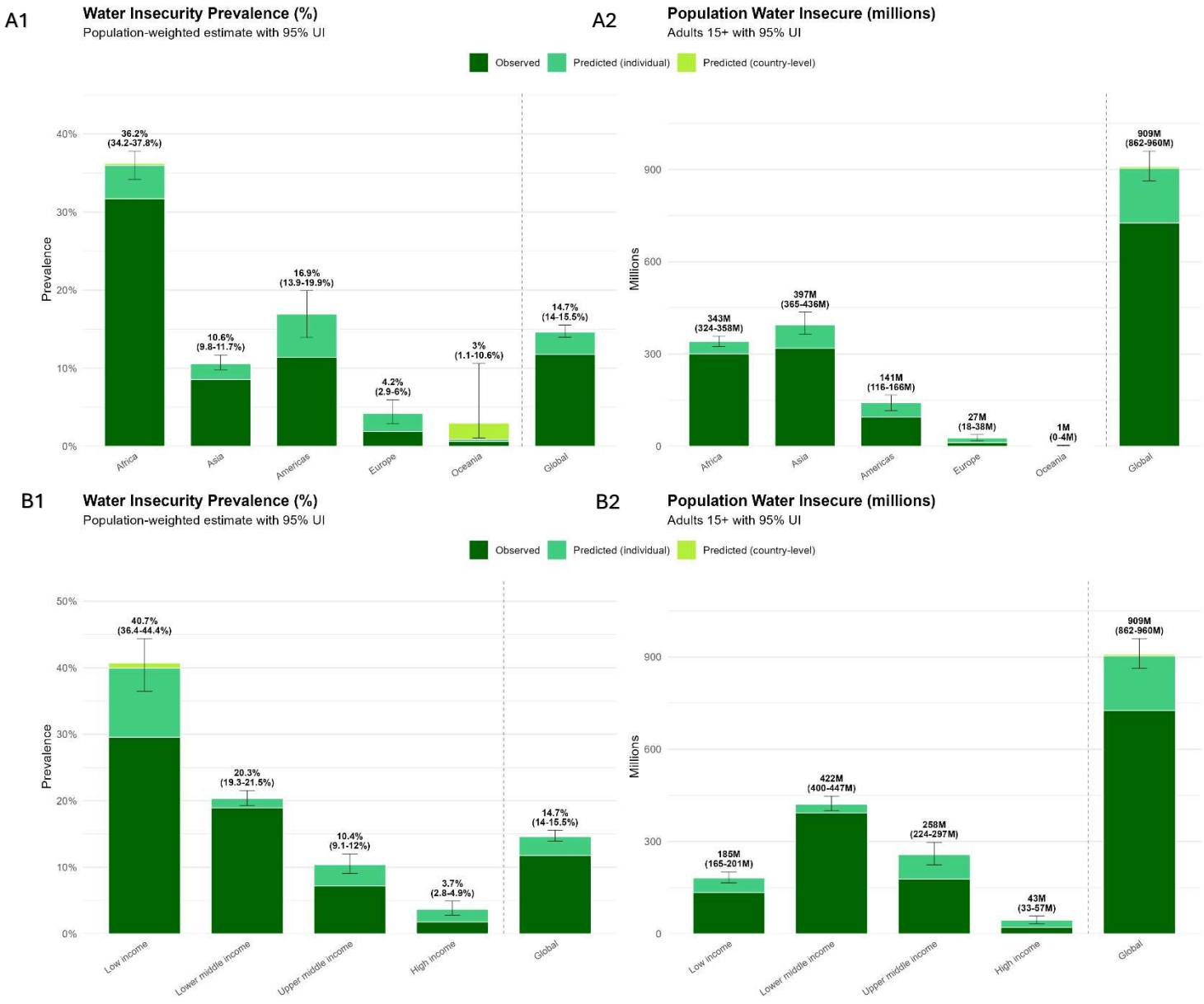
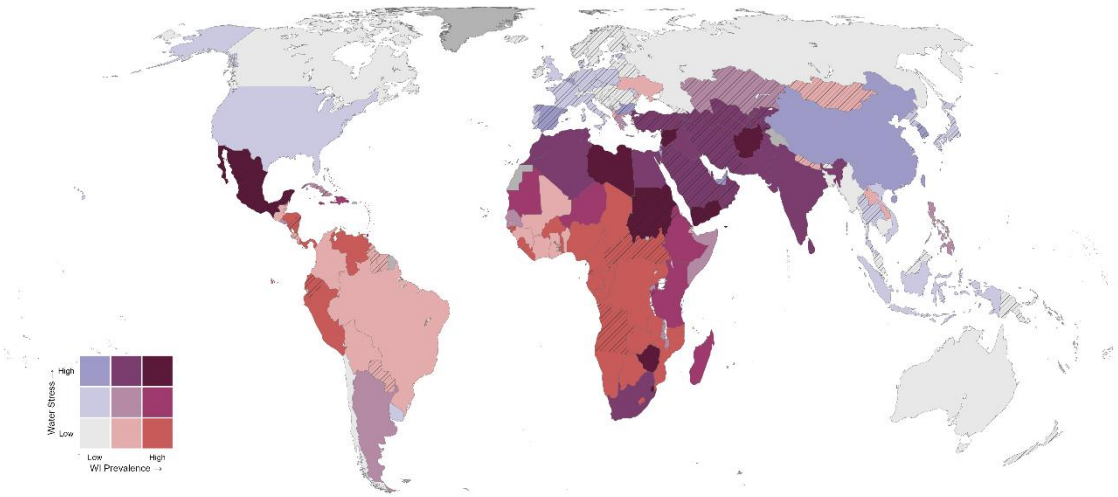
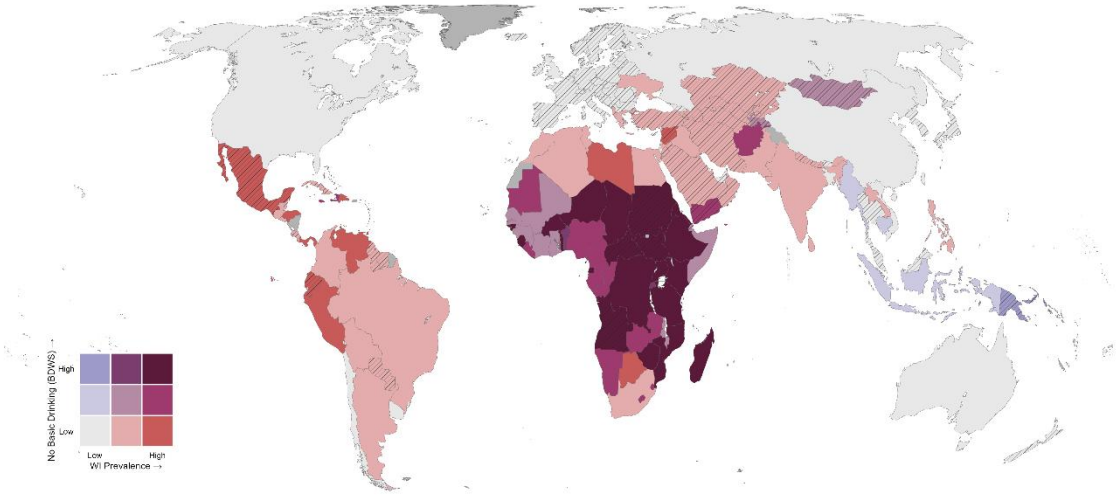


Figure 3: Bivariate maps* showing the relationship between Water Insecurity Prevalence and: A) Level of Water Stress^; B) Prevalence of Population Without Access to At Least Basic Drinking-Water Services^^; C) Prevalence of Population Without Access to Safely Managed Drinking-Water Services^^

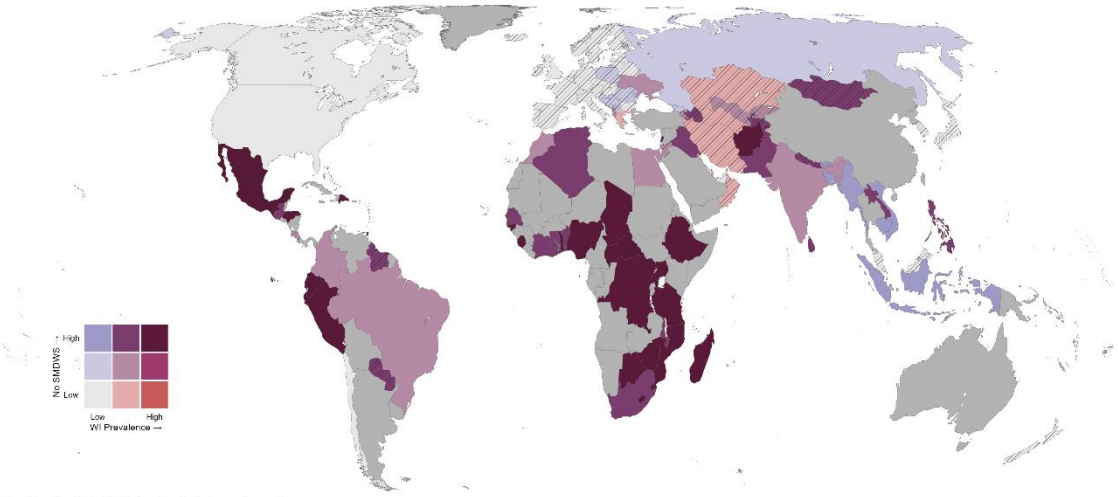
A) Water insecurity prevalence × Water stress



B) Water insecurity prevalence × Lack of basic drinking water access

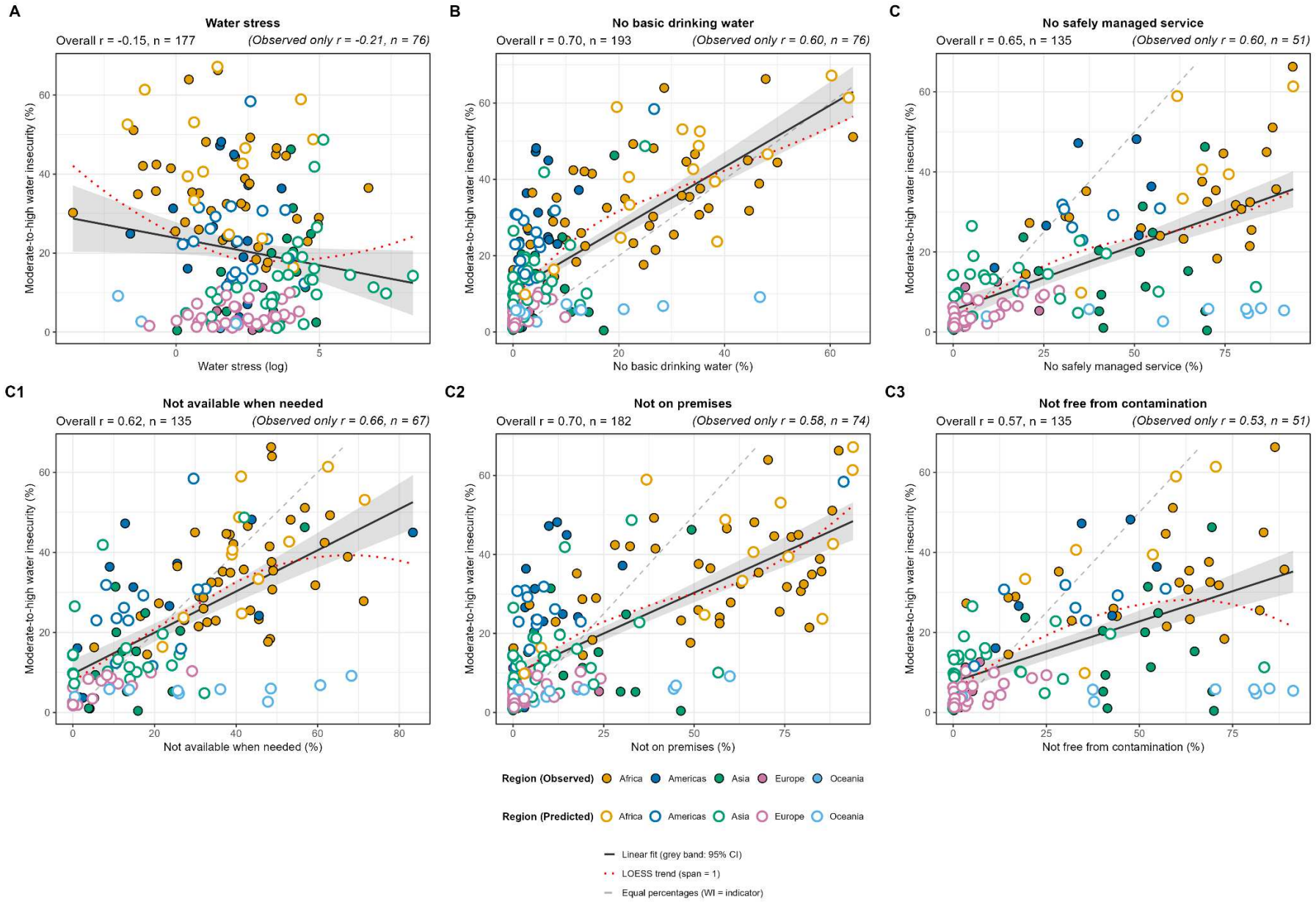


C) Water insecurity prevalence × Lack of access to safely managed drinking water



**Note countries shaded in dark grey indicates no data available. Hatched indicates predicted water insecurity prevalence*
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the authors concerning the legal status of any country, territory, city or area or of its authorities, or concerning, the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which, there may not yet be full agreement.
Map boundaries: UN Geospatial; Abyei, Kosovo and lakes: Natural Earth. Projection: Equal Earth.
^ freshwater withdrawal as a proportion of available freshwater resources, Source: AQUASTAT 2022 via World Bank database [53]
^^ Source: UNICEF/WHO JMP 2024[16]

Figure 4: Correlations between Water Insecurity Prevalence with: A) Level of Water Stress^; B) Prevalence of Population Without Access to At Least Basic Drinking-Water Services^^; C) Prevalence of Population Without Access to Safely Managed Drinking-Water Services^^; C1) Prevalence of population without water available when needed; C2) Prevalence of population without water on premises; C3) Prevalence of population without water free from contamination



^ freshwater withdrawal as a proportion of available freshwater resources, Source: AQUASTAT 2022 via World Bank database [53]
^^ Source: UNICEF/WHO JMP 2024 [16]

Figure 5: Machine Learning Model Validation (Elastic Net) for Predicting Water Insecurity. Results from leave-countries-out-cross-validation (LCO CV). *Panel A: Predicted vs Observed Country Prevalence by region; Panel B: Graph of mean observed vs mean predicted country prevalence by decile of predicted prevalence; Panel C: Predicted- Observed Country Prevalence by Region; Panel D: Source of water insecurity estimate (observed, predicted (individual-level) or predicted (country-level) broken by proportion of countries (N) per region and Total Adult Population (15+) per region*

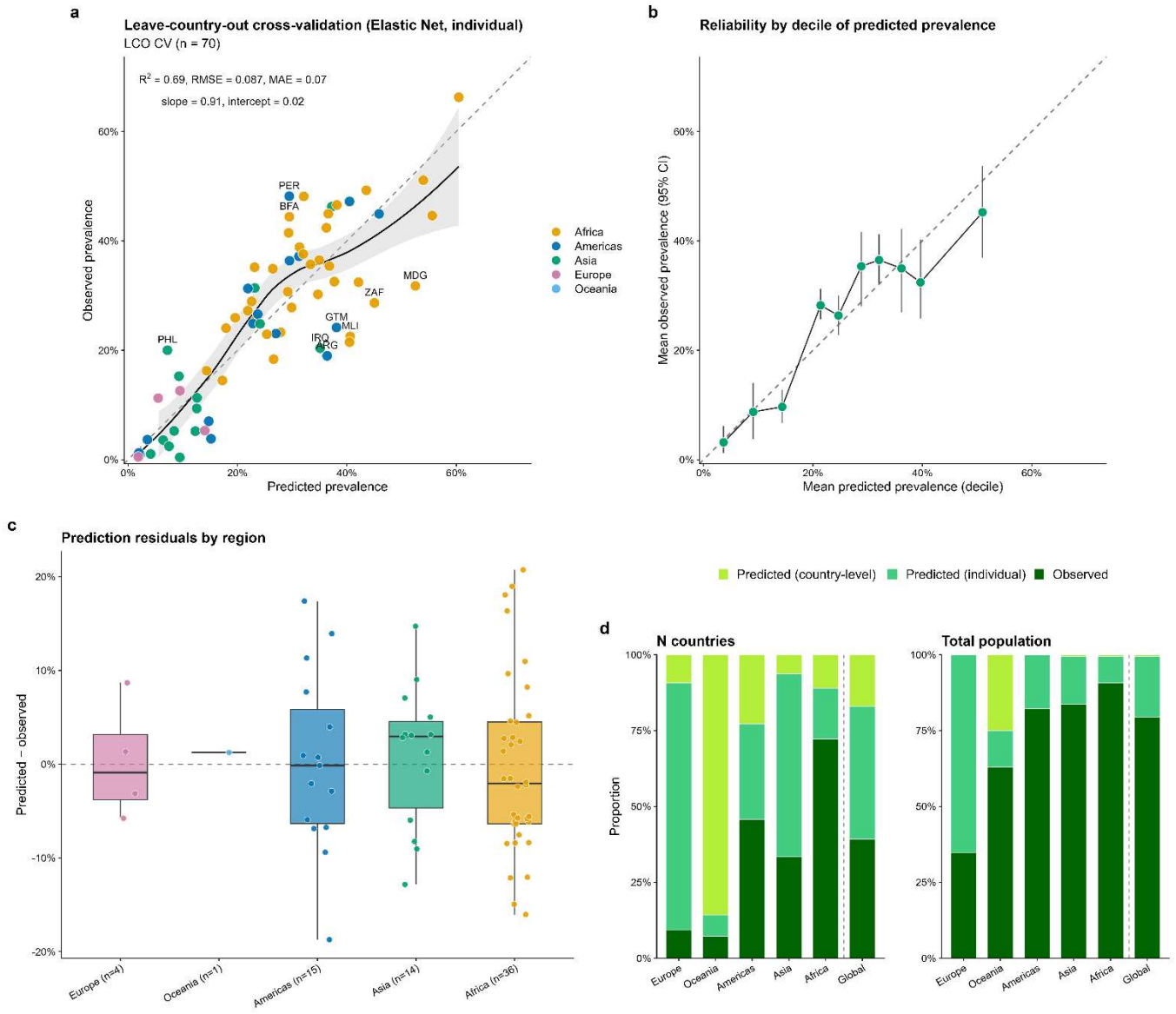
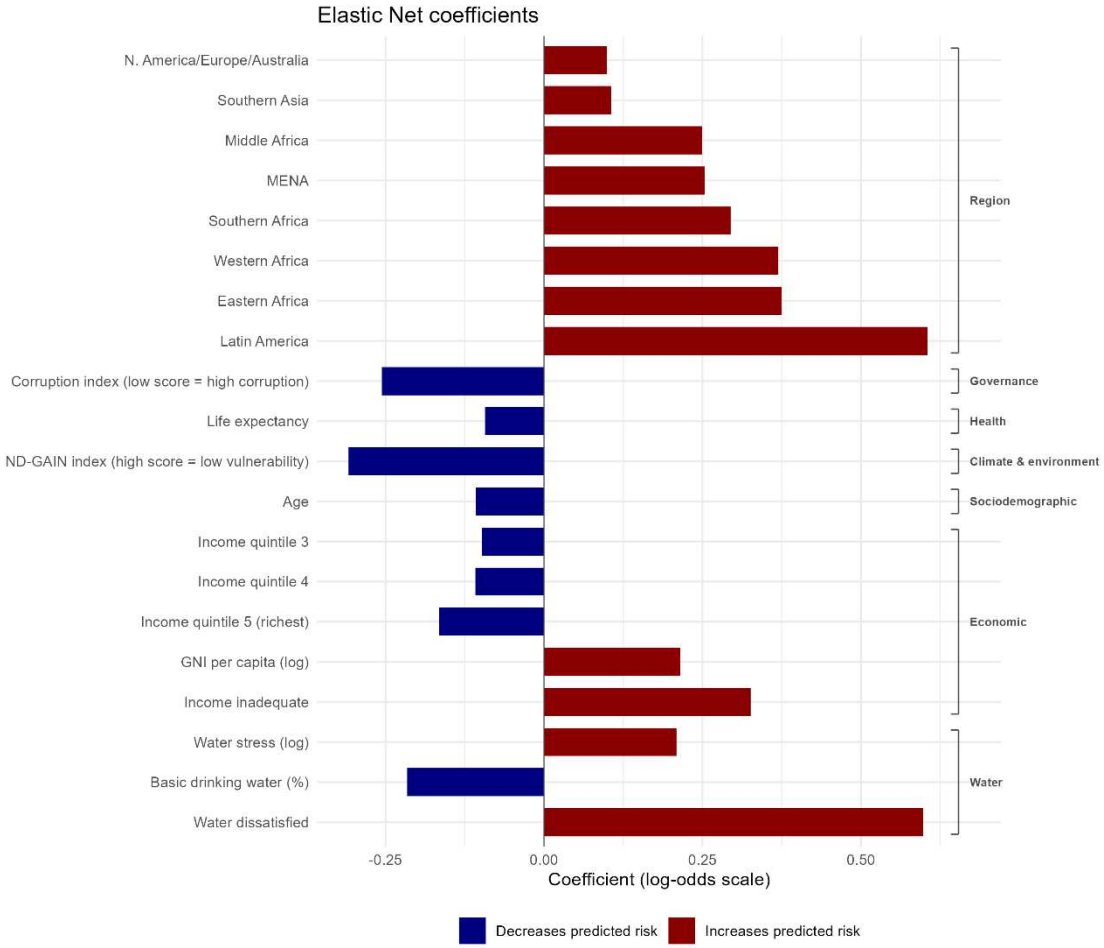


Figure 6: Elastic Net Coefficients-*Final Individual Level Predictive Model (red bars depict predictors that are associated with increased likelihood of being water insecure and blue bars depict predictors that are associated with reduced likelihood of being water insecure)*



4. Methods

4.1 Overview of data and analysis

4.1.1 Data Sources Overview

We analysed individual-survey data on water insecurity experiences from 89,740 respondents, collected through the 2020, 2022, and 2025 Gallup World Polls (GWPs). These surveys provided nationally representative water insecurity (WISE) data across 76 countries. In addition, these surveys collected key sociodemographic indicators for these individuals such as age, gender and economic status (*cf. 4.4 Individual Predictors*).

We also obtained socio-demographic data collected by the GWP that provided nationally representative data for an additional 85 countries (82,805 respondents). For these countries, IWISE data were not captured in the surveys, but the key sociodemographic indicators were captured (*cf. 4.4 Individual Predictors and Table S3*). For each of these countries the latest available survey data were used (the majority were from 2025 n=61 (72%), data ranged from 2006 to 2025).

For all countries (n=194) we obtained national-level predictors, such as level of water stress, perceived level of corruption, access to basic drinking-water services, and GNI per capita from openly accessible databases, such as the World Bank database (cf. *4.5 National-Level predictors and Table S2 for further information on the indicators and sources)*.

4.1.2 Data Analysis Overview

To be able to generate global and regional predictions we used a hierarchical machine learning method to predict estimates for the unobserved countries (n=118). For the total global population, we included 194 countries (i.e., the 193 official UN member countries and the observer state of Palestine).

Among the 118 countries without water insecurity data, 85 had other individual-level socio-demographic survey data collected through GWP (*cf. 4.4 Individual Predictors*). For these countries, machine learning models were tested that used both individual (e.g. age) and national (e.g. GNI) predictors, referred to in this study as the "individual-level model". The model with the best fit was used to predict each individual's probability of experiencing water insecurity, which was then aggregated to estimate country-level prevalences for these 85 countries.

For the additional 33 countries where individual GWP survey data were not available, models were tested that used national-level statistics alone, and the model with the best fit was used to predict the prevalence for these remaining countries, referred to in this study as the "country-level model". The observed estimates and predicted estimates from the individual-level and country-level models were then aggregated to generate global and regional estimates of the prevalence of water insecurity.

4.2 Gallup World Poll (GWP) survey procedures

GWP annually collects nationally representative data on multiple sociodemographic metrics for youth and adults, aged 15 and over, from 161 countries (note some countries are not collected annually so the latest available year was used and any data pre-2015 was flagged, country n= 7). Details of GWP survey methods and sampling frame are described elsewhere [21, 54], see section 4.7.7 for details on how survey weighting was applied to this analysis.

GWP obtained approval for their survey procedures from national governing bodies as required in each country. GWP also obtained informed consent from all participants. For this analysis we received deidentified data from GWP.

4.3 Individual Water Insecurity Experiences (IWISE)

Water insecurity was measured at the individual level through the GWP using the Individual Water Insecurity Experiences (IWISE) Scale, which has been established as valid and cross-country equivalent [22]. The scale consists of 12 questions about the frequency with which individuals have experienced issues with water, including having to limit water-related behaviours (e.g. unable to wash hands or body, change food behaviours), psychosocial effects (e.g. feeling anger or shame related to water issues), and interruptions in water supply. Respondents were asked about the frequency with which they have experienced these issues over the prior 12 months: "never (score 0)", "rarely (1-2 months, score 1)", "sometimes (in some but not every month, score 2)" or "often/always (in almost every month, score 3)". Responses were summed to create a total score ranging from 0-36. For respondents missing data on <4 items, missing item responses were imputed using single imputation (per country by year of data collection). Any respondents missing 4 or more items were dropped from the analysis (n=426, 0.5%).

Moderate-to-high water insecurity was classified as those with IWISE score $\geq$12 [55], referred to throughout the manuscript as "water insecurity". We present this category of water insecurity in this study, rather than any water insecurity, due to the association of multiple key health metrics with moderate-to-high water insecurity [5, 37].

4.4 Individual predictors in GWP

Individual sociodemographic metrics in GWP were identified that may be predictors of water insecurity based on evidence from prior research [17]: 1) age, 2) number of household members, 3) gender, 4) income quintile (produced by GWP), 5) subjective adequacy of income, 6) education, 7) marital status, 8) employment, 9) satisfaction with water quality, and 10) urbanicity (*see Table S3*).

Degree of urbanicity was classified by GWP [56]. When such data were unavailable, a self-reported measure of urbanicity was substituted prior to imputation cf 4.7.1. A variable recording the source of the urbanicity data (classified, self-reported or missing) was included as a

predictor, as were indicators flagging missing urbanicity and missing income quintile, to account for possible bias from missing data.

4.5 National-Level predictors

National statistics from 194 countries (193 UN member States and the observer State of Palestine) were extracted that were plausibly associated with water insecurity experiences. These were identified based on variables found to have importance for predicting anticipated harm from drinking water [57] and malnutrition [58], as well as through expert consultation. The most recent available data were extracted for each country; if countries had data only available before 2015 the data for that predictor was classified as missing. Full list of data extracted is available in *Table S2*, data were downloaded in October 2026 (all data from World Bank was the latest update as of 14/10/2026).

A total of 14 metrics were extracted, three of which were ultimately excluded due to the high proportion of missing data (>20 countries, 10%): food security, primary school completion rate and access to safely managed drinking-water services (SMDWS). GNI per capita, water stress and population density were log transformed, as data were highly skewed.

4.6 Regional classification

For analytical purposes (*c.f. 4.7 Statistical Methods*) countries were classified into nine regions: 1) East & South-East Asia; 2) Southern Asia; 3) Eastern Africa; 4) Middle Africa; 5) Southern Africa; 6) Western Africa; 7) Middle East and North Africa (MENA); 8) Latin America; and 9) North America, Europe & Australia. For the purposes of the machine learning models, these regions were included as a categorical predictor to account for potential systematic between-region differences that may not be fully explained by individual and country-level predictors. These country groupings were derived from UN M49 classifications [59], aggregated to reflect shared aridity and geographic proximity; where sub-regions were merged, countries of broadly similar development context were grouped together. The classifications for each country are displayed in *Table S12* labelled “Region model”.

For regional reporting, countries were grouped according to the official regional grouping used by the UN Statistics Division (M49): 1) Africa; 2)Americas; 3)Asia; 4) Europe; and 5)Oceania, and their sub-regions [59].

4.7 Statistical methods

All data were modelled in R- 4.5.1 using the following key packages: caret [60], glmnet [61], ranger [62], xgboost [63], lme4 [64], betareg [65] mice [66], survey [67], recipes [68] and shapviz[69].

For both the individual-level and country-level analyses, four models were fitted and compared to determine the best-performing model. For the individual-level analysis, three machine

learning models (Elastic Net, Random Forest and XGBoost) were compared with a mixed-effects logistic regression, which allows for unexplained differences between countries. For the country-level analysis, two machine learning models (Elastic Net and XGBoost) were compared with two statistical models suited to proportions (beta regression and fractional logit regression, i.e. a quasi-binomial generalised linear model with a logit link). Models were chosen to suit the format of the outcome in each approach (individual binary outcome; country-level prevalence) and to compare linear (Elastic Net and regression models) and non-linear (Random Forest and XGBoost) approaches. The final models were selected based on their validation performance (Section 4.7.4). Further details on the models are given in Sections 4.7.2 and 4.7.3.

4.7.1 Data Imputations

To fit machine learning models, such as Elastic Net and Random Forests, missing data on predictors had to be imputed. Two main methods of imputation were tested: single imputation and multiple imputation using chained equations [66, 70]. The proportion missing for each predictor was low: <2.3% for individual-level variables and <8.8% for country-level variables (*Tables S2-3*)

For the multiple imputation method (MICE), a two-level strategy was implemented: one for individual-level variables (e.g., age, household size) and one for country-level variables. Imputation methods were chosen by variable type: predictive mean matching for continuous variables, polytomous logistic regression for multi-level categorical variables, and logistic regression for binary variables. Income quintiles were imputed as a numeric variable using predictive mean matching, so that imputed values were always observed quintiles. For country-level imputation, four variables were imputed (Corruption, WASH attributed mortality, Climate Vulnerability Index, log GNI per capita) using seven other country-level indicators as predictors (log water stress, log population density, basic water coverage, life expectancy, fertility rate, under-5 mortality, precipitation). These seven predictors were not imputed by MICE. Four had missing values (water stress (n=17); precipitation (n=13); basic drinking water coverage (n=1); under-5 mortality (n=1)); these were left missing when used as predictors and were replaced afterwards with the median of the country's model region. For individual-level imputation, nine variables were imputed (age, income quintile, inadequate income, education, water dissatisfaction, marital status, urbanicity, employment, household size) using the other individual-level variables as predictors. Region (nine categories; see Section 4.6) was included as a predictor in both levels. For individual-level imputation, region was used in place of country, as binary indicators for each of the 161 countries caused convergence failures. The outcome variable (IWISE) was excluded from the predictor matrix to prevent data leakage. Five imputations were generated with 10 iterations each. For country-level indicators, the mean across the five imputed datasets was used; for individual-level variables, the first completed dataset was used, as most of these variables were categorical and could not be averaged. A single completed dataset was fitted to models, as pooling parameter estimates do not directly extend to machine learning models. Where MICE could not impute a national-level variable as one of the predictors were missing (12 countries, mainly Pacific island states and small European states), the model-region median was used instead. In four countries Cuba, Djibouti,

Guyana and Qatar, at least one individual-level variable was fully missing so MICE imputed these from respondents in other countries. These are flagged in Table S11 as having higher potential uncertainty. Model performance was similar when the models were refitted on each of five MICE imputations (using a simplified imputation model restricted to the 76 observed countries, *Table S14*). A sensitivity analysis was also conducted to test an alternative method for imputing national-level variable for countries missing predictor variables (Supplementary *Text S1; Table S15-S16*). In an additional sensitivity analysis, the predictions from the individual-level model were recalculated, and the model refitted, using each of the five individual-level imputations (*Supplementary Text S2; Table S17*). These did not substantially alter the results.

For the single imputation the following method was used. For individual-level variables (e.g., age, household size) imputations were made using country medians for continuous variables or country mode for categorical variables, except income quintiles. Income quintile was imputed as a numeric median then rounded to the nearest integer. If national-level data (e.g. GNI, Corruption Index) were not available, then data were imputed using regional averages. Similarly for individual-level variables if data were not available at the country level, then regional medians or modes were used. The regions used for these imputations were the nine regions described in section 4.6. As a further sensitivity analysis, income quintile was imputed using a model-based method. In this method for countries with some income data available, an ordinal logistic regression of income quintile on inadequate income, education, urbanicity and country was fitted, and each missing quintile was drawn from the respondent's predicted probabilities. For countries with no income data, the model was fitted on the eight countries with the closest GNI per capita, and respondents were ranked within the country on their predicted value plus random noise and divided into quintiles.

The sensitivity analysis indicated the method of imputation did not affect the model performance (*Table S14*), likely due to the small proportion of missing data. The MICE imputations better captured the distributions of the imputed variables, however, and were therefore used in the final model.

4.7.2 Individual-level machine learning methods

For the individual-level machine learning models, a total of 25 candidate predictors were used (i.e., 10 individual variables, 3 variables flagging missingness of individual urbanicity and income and source of urbanicity data,1 region variable; 11 national-level variables; see *Table S2-4* for full list).

Three machine learning models were fitted to compare linear and nonlinear assumptions: Elastic Net Logistic Regression, XGBoost, and Random Forests. Elastic Net assumes a linear relationship, whilst applying a combined lasso and ridge penalty to prevent overfitting; variables are selected automatically by shrinking weak predictors towards zero. Random Forests and extreme gradient boosting (XGBoost) are both tree-based ensemble models, with Random Forests aggregating predictions from multiple independently grown decision trees and XGBoost sequentially fitting new trees that correct the residual errors of the existing ensemble. A mixed-

effects logistic regression was also fitted to assess whether the measured predictors adequately captured between country variation.

4.7.3 Country-level machine learning methods

For the country-level machine learning models, a total of 12 predictors were used (the same 11 national-level variables and 1 region variable; see *Table S2 & 4* for full list). For these models, the outcome variable was defined as the proportion of water insecurity in each country. Four models were fitted: two using traditional statistical models (beta regression and fractional logit regression, i.e. a quasi-binomial generalised linear model with a logit link) and two machine learning models (Elastic Net and XGBoost). The traditional models are designed to model proportional outcomes without regularisation, with quasibinomial models being simpler than beta-regression.

4.7.4 Model Validation Methods

To validate the machine learning models, six countries were held out to create a test data set, leaving 70 countries in the training data. India was held out, as this was the most sampled country and could potentially skew the respondent-level performance metrics. An additional five countries were randomly held out (Namibia, Somalia, Palestine, Brazil and Cameroon). These six countries were not used during model training to be able to discern if the model was able to generalise to entirely unseen country contexts. The final individual-level model was fitted to the 70 training countries, but as the country-level model only had 76 observations the model was refitted to all 76 observed countries. The ROC and AUC were compared for each model of the hold-out countries to determine model fit.

Furthermore, leave-countries-out cross-validation (LCO CV) was conducted, where within the training dataset countries were grouped randomly into five folds (all observations for each country were assigned to the same fold). After running these folds, predictions were generated for each observation, which were averaged using survey weights to generate predicted country estimates (only for individual-level models) and were compared to the observed country estimates. To further assess model fit using the LCO CV predictions, we calculated the R^2 (squared Pearson correlation), which measures how well the model captures between-country variation in prevalence, and MAE (Mean absolute Error), which quantifies the absolute average difference in predictions compared to observed prevalences. Both methods were conducted for validation purposes as the hold-out countries provide an independent test of the models, whereas LCO CV gives more stable performance estimates across a larger number of countries; LCO CV results are therefore used for model comparison, and hold-out results are also reported (*Table S5-8*).

For the individual-level model to calculate classification metrics (accuracy, sensitivity, specificity, F1), we classified individuals as experiencing moderate-to-high water insecurity if their predicted probability was at or above the threshold that maximised Youden's index (sensitivity + specificity − 1) in each evaluation set (*Tables S5-S6*).

In addition to examining model fit, model calibration for the final individual- and country-level models were assessed using the LCO CV predictions, as this has been found to be critical step to validate models in predictive analytics [71]. The observed country prevalences were regressed on the predictive estimates using linear regression. The slope and the intercept of these regressions were assessed, with ideal calibration corresponding to an intercept of 0 and a slope of 1.

For the individual model, from the country-hold validation and the LCO CV, the Elastic Net Model performed the best (*Table S5-7; Figure S8-12).* In a mixed-effects model with a country-specific random intercept, a subset of 13 predictors and region explained 79% of the between-country variance (*Table S6 footnote*), supporting the use of these predictors to estimate prevalence in countries without data. For the country-level model, Elastic Net was also found to have the highest cross-validated performance from the LCO CV (*Table S7-8, Figure S11-14*).

4.7.5 Final Individual and Country Level Model

For both the individual and country level models, Elastic Net had the best fit. The individual-level models, however, had a better fit overall than the country-level models (*Table S7*), so these predictions were preferred for countries where data were available.

For the individual-level models, the probability of being water insecure was estimated for all individuals within each country, these probabilities were then averaged using survey-weights (*see 4.7.7 for details on these weights*) to estimate national prevalence. For the country-level models an overall national predicted prevalence was estimated for each country.

Feature selection and SHAP (SHapley Additive exPlanations) values [72] were used to understand the importance of different predictors for estimating water insecurity. Both Elastic-Net coefficients and SHAP feature importance from XGBoost were plotted and compared to understand how predictors compared between linear and gradient-boosted decision-tree approaches (see **Figure 6** and *Figure S15-19*).

Estimates were flagged as potentially uncertain if the value of the predictors fell beyond the values included in the training data by ≥10% (e.g. if the value of water stress for a country exceeded the maximum value in the training dataset) or if there was much disagreement between models (Elastic Net vs. XGBoost), or for countries with individual-level data, large disagreement between the country-level vs. individual-level estimates. Countries whose GWP data was collected before 2015 were also flagged (see *Table S11-12*).

We also compared modelled estimates where possible with collected nationally survey data from other sources, see Supplemental Text S3.

4.7.6 Sensitivity Analysis

As satisfaction with water quality is a subjective indicator that measures a related, though not identical construct to water insecurity, sensitivity analysis was run excluding this predictor from

the individual-level models to understand how this indicator affected feature selection and model performance.

We conducted a further sensitivity analysis using alternative regional model groupings for five countries. Japan, the Republic of Korea and Singapore were re-classified to North America, Europe & Australia, reflecting their closer similarity to other high-income countries, and Sudan and Yemen to Eastern Africa, reflecting their closer similarity to neighbouring countries in the Horn of Africa than to the rest of MENA. As none of these countries were in the observed dataset, the reclassification did not alter the training data or the fitted model. Estimates changed by 0.4 to 0.8 percentage points for the three East Asian countries and by up to 5.3 points for Sudan and Yemen (*Table S13*); given estimated prevalences above 48% in the latter two, these shifts did not substantially alter their classification as settings with relatively higher prevalence settings or any substantive conclusion.

4.7.7 Survey-weighting for country, regional and global estimates

GWP contains weights that adjust for survey methods to ensure national representativeness (further details are provided in [21, 54]). These weights were normalised by dividing each observation's weight by the total number of observations in its country, so that in pooled analysis each countries' representativeness was preserved whilst preventing countries with larger samples from dominating. For regional, income-group and global estimates, projection weights were calculated by multiplying the normalised weights by each country's estimated 2025 population of youth and adults 15 years and over, using data obtained from the World Bank Population database. This method ensured that population size was accounted in global, regional, sub-regional and income-group estimates. For countries where only estimates from the country-level model were available (no individual GWP survey data), country prevalence estimates were weighted by the youth and adult (≥15 years) population size. Country income groups followed the World Bank FY2027 classification.

To calculate the uncertainty around the estimates, 95% uncertainty intervals (UIs) were estimated using methods appropriate for the way in which the country prevalences were estimated. For the observed countries, uncertainty intervals were derived from the survey design using Taylor series linearisation (accounting for primary sampling units, strata and sampling weights) to reflect sampling uncertainty, with intervals calculated on the logit scale. For predicted countries and for regional, sub-regional, income group and global estimates, UIs were obtained from a bootstrap with 1,000 repetitions that also captures model uncertainty. In each repetition, the training countries were randomly resampled with replacement within each model region (with primary sampling units also resampled within countries for the individual-level model, and the observed prevalences sampled from their survey distributions for the country-level model), and both Elastic Net models were refitted. Each predicted country's prevalence was recalculated from the refitted model (for countries estimated with the individual-level model, primary sampling units were also resampled to reflect sampling uncertainty in the underlying individual-level data), and a residual drawn from that model's out-of-sample errors in the observed countries was added on the logit scale to reflect country-specific prediction error. The logit scale was used to keep values between 0 and 100% without

truncation. For the regional and global estimates, observed countries' prevalences were drawn in each repetition from normal distributions based on their survey estimates and standard errors. Regional, sub-regional, income group and global prevalences were calculated in each repetition as population-weighted means of all 194 countries. UIs are the 2.5th and 97.5th percentiles across repetitions; point estimates are those from the observed data and the final models. Results from alternative methods (a residual bootstrap, resampling countries within groups, and treating country errors as independent) are presented in *Figure S20*. The UIs do not capture systematic differences between observed and unobserved countries or imputation uncertainty.

4.7.8 Bivariate Mapping and correlations with other water-related indicators

To examine how estimates of water insecurity relate with other national indicators related to water we produced four bivariate choropleth maps, each displaying water insecurity prevalence against one of: 1) water stress (freshwater withdrawal as a proportion of available freshwater resources); 2) the proportion of the population without access to at least basic drinking-water services; and 3) the proportion of the population without access to safely managed drinking-water services (SMDWS); and 4) drinking water not availability when needed (a component of SMDWS). Only non-imputed data for these indicators were used in this analysis. Each variable was classified into three categories (water insecurity prevalence: <10%, 10–<30%, ≥30%; populations without basic, safely managed or available drinking water: <10%, 10–<30%, ≥30%; water stress: <10%, 10–<40%, ≥40%), and the two variables were combined into a 3×3 colour matrix, so that the colour displayed on the map represents the joint level on both variables.

In addition, we ran correlations and partial correlations, adjusted for GNI, between water insecurity prevalence (using both predicted and observed data, and observed data only) with the four indicators listed above as well as the other two components of SMDWS: 5) water not available on premises; 6) not free from contamination. We ran partial correlations by first regressing each variable on log-transformed GNI and then plotting the residuals from these regressions in separate graphs. In each graph the residuals of the water insecurity regression was on the y-axis and the residuals from the regression of each water indicator was on the x-axis. A fitted linear line and loess line (span 1 to provide a smooth trend) was included in each plot to visualise these correlations.

4.8 Ethics

This study is based on deidentified survey collected by GWP, and national-level data that is publicly available. GWP obtained informed consent from all participants and obtained ethical approvals from Gallup's Internal Review Board and by the governing bodies in countries where approval is required. The authors were not involved in the data collection process.

Acknowledgements:

This work is funded through the ANH Academy Science-Policy Platform, led by the London School of Hygiene & Tropical Medicine (LSHTM), in partnership with Tufts University and the University of Sheffield. ANH Academy is co-funded by the UK International Development from the UK government and by the Gates Foundation 400196-401/INV-042668. The authors also acknowledge the support from Elizabeth Walder, the Karin Bain and John Kukral Foundation, support of the American people provided to the Feed the Future Sustainable Intensification Innovation Lab through the United States Agency for International Development Cooperative Agreement (AID-OAA-L-14-00006) and Northwestern University, including from the Northwestern Institute on Complex Systems, the Global Poverty Research Lab, Weinberg College, the Office for Research, Leslie and Mac McQuown via the Center for Engineering Sustainability and Resilience, the Buffett Institute for Global Affairs, the Center for Water Research, the Israel Innovation Project, and the Paula M. Trienens Institute for Sustainability and Energy. SLY was supported by a Leverhulme Trust Visiting Professorship, the Carnegie Foundation, and the Guggenheim Foundation. Finally, we gratefully acknowledge the important contributions of Hilary Bethancourt towards cleaning and collating the 2020 and 2022 GWP IWISE data, and Adelaide Dalton for her support in verifying external data sources. The funders had no role in study design, data analysis, decision to publish, or preparation of the manuscript. The conclusions and opinions expressed in this work are those of the authors alone and shall not be attributed to any of the funders.

Author Contributions

IB, SLY, EAF and JDM conceptualised the study. IB led the formal analysis, JDM, EAF, EEG and PM contributed to the analysis and methods. SLY, EAF and SK supervised. IB led the drafting of the manuscript, all authors (SLY, JDM, EAF, EEG, PM, KG, SK and RJ) contributed to the overall writing, reviewing and editing of the manuscript. SLY and SK secured funding for the study.

Declaration of Interests

The authors declare no competing interests

Data and Code Availability

The R code for the analysis is openly available at: IEBose/Global-Estimates-of-Water-Insecurity-: R scripts for the global estimates of water insecurity. The details of how to access the national-level data that are publicly available is also detailed in the READ_ME file; a copy of the data used for this analysis is also available on the github repository. The full deidentified GWP dataset and data dictionary can be requested from GallupHelp@mail.gallup.com

References

1. United Nations General Assembly. Resolution 64/292: The human right to water and sanitation. New York: 2020 03/08/2010. Report No.: A/RES/64/292.
2. Rosinger AY, Young SL. The toll of household water insecurity on health and human biology: Current understandings and future directions. WIREs Water. 2020;7(6):e1468. doi: https://doi.org/10.1002/wat2.1468.
3. Bose I, Frongillo E, Miller JD, Sparling TM, Mayer A, Viviani S, et al. Does Water Insecurity Mediate the Associations Between Climate Hazards and Food Insecurity? An Analysis of Nationally Representative Data From 23 Countries. Preprint. 2026.
4. Miller JD, Young SL, Bryan E, Ringler C. Water insecurity is associated with greater food insecurity and lower dietary diversity: panel data from sub-Saharan Africa during the COVID-19 pandemic. Food Security. 2024;16(1):149-60. doi: 10.1007/s12571-023-01412-1.
5. Young SL, Bethancourt HJ, Frongillo EA, Viviani S, Cafiero C. Concurrence of water and food insecurities, 25 low-and middle-income countries. Bulletin of the World Health Organization. 2023;101(2):90-101.
6. Ringler C, Agbonlahor M, Barron J, Baye K, Meenakshi J, Mekonnen DK, Uhlenbrook S. The role of water in transforming food systems. Global Food Security. 2022;33:100639.
7. Bose I, Dreibelbis R, Green R, Murray KA, Ceesay O, Kovats S. Climate change, seasonality and household water security in rural Gambia: A qualitative exploration of the complex relationship between weather and water. PLOS Water. 2024;3(6):e0000239.
8. Bose I, Frongillo E, Dooley C, Sparling T, Kadiyala S, Young S. Climate hazards and water insecurity, by geographic and economic vulnerabilities. 2025.
9. He C, Liu Z, Wu J, Pan X, Fang Z, Li J, Bryan BA. Future global urban water scarcity and potential solutions. Nature Communications. 2021;12(1):4667. doi: 10.1038/s41467-021-25026-3.
10. Mekonnen MM, Hoekstra AY. Four billion people facing severe water scarcity. Science Advances. 2016;2(2):e1500323. doi: doi:10.1126/sciadv.1500323.
11. Jepson WE, Wutich A, Colllins SM, Boateng GO, Young SL. Progress in household water insecurity metrics: a cross-disciplinary approach. WIREs Water. 2017;4(3):e1214. doi: https://doi.org/10.1002/wat2.1214.
12. Liu J, Yang H, Gosling SN, Kummu M, Flörke M, Pfister S, et al. Water scarcity assessments in the past, present, and future. Earth's Future. 2017;5(6):545-59. doi: https://doi.org/10.1002/2016EF000518.
13. Octavianti T, Staddon C. A review of 80 assessment tools measuring water security. WIREs Water. 2021;8(3):e1516. doi: https://doi.org/10.1002/wat2.1516.
14. UN Department of Economic and Social Affairs. Ensure availability and sustainable management of water and sanitation for all [cited 2026 19th October]. Available from: https://sdgs.un.org/goals/goal6#overview.
15. FAO & UN-Water. Progress on the level of water stress – Mid-term status of SDG Indicator 6.4.2 and acceleration needs, with special focus on food security. Rome: FAO, 2024 9251390681.
16. UNICEF, WHO. Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) https://washdata.org/ [cited 2026 20/07].
17. Young SL, Bethancourt HJ, Ritter ZR, Frongillo EA. Estimating national, demographic, and socioeconomic disparities in water insecurity experiences in low-income and middle-income countries in 2020–21: a cross-sectional, observational study using nationally representative survey data. The Lancet Planetary Health. 2022;6(11):e880-e91.
18. Bose I, Dreibelbis R, Green R, Murray KA, Ceesay O, Kovats S. Coping strategies for household water insecurity in rural Gambia, mediating factors in the relationship between weather, water and health. BMC Public Health. 2024;24(1):3150.

19. Collins SM, Mbullo Owuor P, Miller JD, Boateng GO, Wekesa P, Onono M, Young SL. 'I know how stressful it is to lack water!'Exploring the lived experiences of household water insecurity among pregnant and postpartum women in western Kenya. Global Public Health. 2019;14(5):649-62.
20. Wutich A, Brewis A, Tsai A. Water and mental health. Wiley Interdisciplinary Reviews: Water. 2020;7(5):e1461.
21. Young SL, Bethancourt HJ, Ritter ZR, Frongillo EA. The Individual Water Insecurity Experiences (IWISE) Scale: reliability, equivalence and validity of an individual-level measure of water security. BMJ Global Health. 2021;6(10):e006460. doi: 10.1136/bmjgh-2021-006460.
22. Miller JD, Bose I, Frongillo E, Young SL. Validation of the Individual Water Insecurity Experiences Scale in 76 Countries to Assess Prevalence and Disparities. Available at SSRN 7165562.
23. Young S, Miller J, Bose I. Measuring human experiences to advance safe water for all. Evanston: Institute for Policy Research, Northwestern University. 2024.
24. Shemer H, Wald S, Semiat R. Challenges and Solutions for Global Water Scarcity. Membranes. 2023;13(6):612. PubMed PMID: doi:10.3390/membranes13060612.
25. Van Vliet MT, Jones ER, Flörke M, Franssen WH, Hanasaki N, Wada Y, Yearsley JR. Global water scarcity including surface water quality and expansions of clean water technologies. Environmental Research Letters. 2021;16(2):024020.
26. UN Water. SDG 6 Country Acceleration Case Studies 2023 Singapore. 2023.
27. Jensen O, Nair S. Integrated Urban Water Management and Water Security: A Comparison of Singapore and Hong Kong. Water. 2019;11(4):785. PubMed PMID: doi:10.3390/w11040785.
28. Hernández-Vásquez A, Vargas-Fernández R. Beyond the faucet: social-geographic disparities and trends in intermittent water supply in Peru. Journal of Water and Health. 2025;23(1):1-15.
29. Huberts A, Palma D, Bernal García AC, Cole F, Roberts EFS. Making scarcity "enough": The hidden household costs of adapting to water scarcity in Mexico City. PLOS Water. 2023;2(3):e0000056. doi: 10.1371/journal.pwat.0000056.
30. Artell J, Ahtiainen H, Pouta E. Subjective vs. objective measures in the valuation of water quality. Journal of Environmental Management. 2013;130:288-96. doi: https://doi.org/10.1016/j.jenvman.2013.09.007.
31. Venkataramanan V, Collins SM, Clark KA, Yeam J, Nowakowski VG, Young SL. Coping strategies for individual and household-level water insecurity: A systematic review. WIREs Water. 2020;7(5):e1477. doi: https://doi.org/10.1002/wat2.1477.
32. Achore M, Bisung E, Kuusaana ED. Coping with water insecurity at the household level: A synthesis of qualitative evidence. International Journal of Hygiene and Environmental Health. 2020;230:113598. doi: https://doi.org/10.1016/j.ijheh.2020.113598.
33. Majuru B, Suhrcke M, Hunter PR. How Do Households Respond to Unreliable Water Supplies? A Systematic Review. International Journal of Environmental Research and Public Health. 2016;13(12):1222. PubMed PMID: doi:10.3390/ijerph13121222.
34. Bethancourt HJ, Bose I, Tupou LM, Meinhart M, Frongillo EA, Greaves A, Young SL. Mental Distress and Resource Insecurity in the Kingdom of Tonga After the 2022 Volcanic Eruption: Associations of Water and Food Insecurity With Mental Distress. Journal of the Academy of Nutrition and Dietetics. doi: 10.1016/j.jand.2025.08.002.
35. Young SL, Bethancourt HJ, Frongillo EA, Viviani S, Cafiero C. Concurrence of water and food insecurities, 25 low-and middle-income countries. Bulletin of the World Health Organization. 2022;101(2):90.
36. Roba KT, Roba AA, Hassen JY, Thompson G, Brewis A, Rosinger AY. Household Water Insecurity Is Associated With Greater Likelihood of Early Exclusive Breastfeeding Cessation and Diarrhea Among Children Aged 6 to 59 Months in Ethiopia: Data From Haramaya University

Health and Demographic Surveillance System. Journal of the Academy of Nutrition and Dietetics. 2026;126(6):156303. doi: https://doi.org/10.1016/j.jand.2026.156303.
37. Bose I, Bethancourt HJ, Shamah-Levy T, Mundo-Rosas V, Muñoz-Espinosa A, Ginsberg T, et al. Mental health, water, and food: Relationships between water and food insecurity and probable depression amongst adults in Mexico. Journal of Affective Disorders. 2024. doi: https://doi.org/10.1016/j.jad.2024.10.116.
38. Mack EA, Wrase S. A burgeoning crisis? A nationwide assessment of the geography of water affordability in the United States. PloS one. 2017;12(1):e0169488.
39. Schillinger J, Özerol G, Güven-Griemert Ş, Heldeweg M. Water in war: Understanding the impacts of armed conflict on water resources and their management. Wiley Interdisciplinary Reviews: Water. 2020;7(6):e1480.
40. McLain AC, Frongillo EA, Feng J, Borghi E. Prediction intervals for penalized longitudinal models with multisource summary measures: an application to childhood malnutrition. Statistics in medicine. 2019;38(6):1002-12.
41. Greenwood EE, Lauber T, van den Hoogen J, Donmez A, Bain RE, Johnston R, et al. Mapping safe drinking water use in low-and middle-income countries. Science. 2024;385(6710):784-90.
42. Geere JL, Mokoena MM, Jagals P, Poland F, Hartley S. How do children perceive health to be affected by domestic water carrying? Qualitative findings from a mixed methods study in rural South Africa. Child: Care, Health and Development. 2010;36(6):818-26. doi: https://doi.org/10.1111/j.1365-2214.2010.01098.x.
43. Rhue SJ, Torrico G, Amuzie C, Collins SM, Lemaitre A, Workman CL, et al. The effects of household water insecurity on child health and well-being. WIREs Water. 2023;10(6):e1666. doi: https://doi.org/10.1002/wat2.1666.
44. Rhue S. Acknowledged but Unheard: The Absence of Children's Voices in Water Insecurity. 2021.
45. Frongillo EA, Fram MS, Ghattas H, Bernal J, Jamaluddine Z, Kirkpatrick SI, et al. Development, validity, and cross-context equivalence of the child food insecurity experiences scale for assessing food insecurity of school-age children and adolescents. The Journal of Nutrition. 2022;152(9):2135-44.
46. Rosinger AY, Douglass M, McGrosky A, Roba KT, Davies B, Meriwether N, et al. Historic drought in northern Kenya shifts diet and mobility away from traditional pastoralism as resource security and nutrition falter. Global Environmental Change. 2026;100:103212. doi: https://doi.org/10.1016/j.gloenvcha.2026.103212.
47. Pradhananga N, Enriquez JP, Okronipa H, Holston D, Sadler JM. Dual Burden of Food and Water Insecurity Among SNAP Households with Children in the Southern United States. International Journal of Environmental Research and Public Health. 2026;23(7):891. PubMed PMID: doi:10.3390/ijerph23070891.
48. Young SL, Miller JD, Bose I, Collins SM, Danaj S, Kazakov B, et al. The Water Insecurity Experiences (WISE) Scales are suitable for use in high-income settings: findings from cognitive interviews and nationally representative surveys. International journal for equity in health. 2025;24(1):338.
49. Macharia D, MacDonald L, Mugabo L, Donovan K, Brooks W, Gudissa S, et al. Mixed methods study design, pre-analysis plan, process evaluation and baseline results of trailbridges in rural Rwanda. Science of The Total Environment. 2022;838:156546.
50. Shamah-Levy T, Mundo-Rosas V, Munoz-Espinosa A, Humarán IM-G, Perez-Escamilla R, Melgar-Quinonez H, et al. Feasibility of a water insecurity experiences scale in Mexican households. Salud Pública de México. 2023;65(3):219-26.
51. World Bank. Pasifika Pathways for Water Security. Washington, DC: The World Bank, forthcoming.

52. King Khalid Foundation. Email: WISE data Saudi Arabia. In: Indira Bose and Sera Young, editor. 2026.
53. World Bank Development Indicators. Level of water stress: freshwater withdrawal as a proportion of available freshwater resources, source: FAO AQUASTAT. https://data360.worldbank.org/en/indicator/WB_WDI_ER_H2O_FWST_ZS?country=&view=bar&minYear=1962&maxYear=2022.
54. Gallup World Poll. Worldwide research methodology and codebook. 2026.
55. Frongillo EA, Bethancourt HJ, Miller JD, Young SL. Identifying ordinal categories for the Water Insecurity Experiences Scales. Journal of Water, Sanitation and Hygiene for Development. 2024:washdev2024042.
56. Dijkstra L, Florczyk AJ, Freire S, Kemper T, Melchiorri M, Pesaresi M, Schiavina M. Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. Journal of Urban Economics. 2021;125:103312.
57. Miller JD, Staddon C, Salzberg A, Lucks JB, Bruine de Bruin W, Young SL. Self-reported anticipated harm from drinking water across 141 countries. Nature communications. 2024;15(1):7320.
58. Milman A, Frongillo EA, de Onis M, Hwang J-Y. Differential improvement among countries in child stunting is associated with long-term development and specific interventions. The Journal of Nutrition. 2005;135(6):1415-22.
59. United Nations. Statistical Office. Standard country or area codes for statistical use: UN; 1982.
60. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. Package 'caret'. The R Journal. 2020;223(7):48.
61. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, Qian J. Package 'glmnet'. CRAN R Repositary. 2021;595:874.
62. Wright MN, Wager S, Probst P, Wright MMN. Package 'ranger'. Version 011. 2019;2.
63. Chen T, He T, Benesty M, Khotilovich V. Package 'xgboost'. R version. 2019;90(1-66):40.
64. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. Package 'lme4'. convergence. 2015;12(1):2.
65. Zeileis A, Cribari-Neto F, Gruen B, Kosmidis I, Simas AB, Rocha AV, Zeileis MA. Package 'betareg'. R package. 2016;3(2):51.
66. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. Journal of statistical software. 2011;45:1-67.
67. Lumley T, Lumley MT. Package 'survey'. Available at the following link: https://cran r-project org. 2020.
68. Kuhn M, Wickham H, Hvitfeldt E. recipes: Preprocessing and Feature Engineering Steps for Modeling. 1.3.3 ed2026.
69. Mayer M. shapviz: SHAP Visualizations. 0.10.3 ed2025.
70. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Annals of translational medicine. 2016;4(2):30.
71. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. Epub 20191216. doi: 10.1186/s12916-019-1466-7. PubMed PMID: 31842878; PubMed Central PMCID: PMCPMC6912996.
72. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clinical and Translational Science. 2024;17(11):e70056. doi: https://doi.org/10.1111/cts.70056.

Supplementary Files

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

- SuppMaterialsGlobalWIestimatesv7.pdf