TECHNICAL NOTE N° IDB-TN-03034

Behind the numbers:

A comparative analysis of international datasets on water and sanitation in Latin America and the Caribbean

Author:
Jesse Madden Libra

Editors:
María Eugenia de la Peña
María Pérez Urdiales

Inter-American Development Bank
Water and Sanitation Division

September 2024

Behind the numbers:

A comparative analysis of international datasets on water and sanitation in Latin America and the Caribbean

Author:
Jesse Madden Libra

Editors:
María Eugenia de la Peña
María Pérez Urdiales

Inter-American Development Bank
Water and Sanitation Division

September 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Libra, Jesse.
Behind the numbers: a comparative analysis of international datasets on water and sanitation in Latin
America and the Caribbean / Jesse Madden Libra; editors, María Eugenia de la Peña, María Pérez Urdiales.
p. cm. — (IDB Technical Note ; 3034)

Includes bibliographical references.

1. Water supply-Latin America. 2. Water supply-Caribbean Area. 3. Sanitation-Latin America. 4. Sanitation-Caribbean Area. 5. Data sets. I. De la Peña, María Eugenia, editor. II. Pérez Urdiales, María, editor. III.
Inter-American Development Bank. Water and Sanitation. IV. Title. V. Series.
IDB-TN-3034

JEL Codes: Q25

Keywords: water access, water utilities, water, data, surveys

http://www.iadb.org

IDB

OLAS

Behind the numbers:

A comparative analysis of international datasets on water and sanitation in Latin America and the Caribbean

Author:
Jesse Madden Libra
Editors:
María Eugenia de la Peña, María Pérez Urdiales
September 2024

Table of Contents

Acknowledgements

I would like to extend my gratitude to the reviewers: Santiago Cunial and Ricardo Martinez with the Water and Sanitation Division at the Inter-American Development Bank (IDB), Cristina Pombo Rivera and her team in the Social Division at the IDB, and the team at the WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation (JMP) for their valuable feedback and suggestions. A special thank you is owed to Richard Johnston of the JMP for his in-depth revision and feedback which contributed to the quality of this document.

Thank you to editors Maria Perez Urdiales and Maria Eugenia de la Peña for their support and guidance throughout the development of this document.

Abbreviations

AB-LAPOP	AmericasBarometer Survey from the Latin American Public Opinion Project
CASEN	Chile's national household survey, the Encuesta de Caracterización Socioeconómica Nacional
CSSP	Trinidad and Tobago's national household survey, the Continuous Sample Survey of Population
DHS	Demographic and Health Surveys
ECH	Bolivia's national household survey, the Encuesta de Hogares
ECH	Uruguay's national household survey, the Encuesta Continua de Hogares
ECVMAS	Haiti's national household survey, the Enquête sur les Conditions de Vie des Ménages après Séisme
EHM	Venezuela's national household survey, the Encuesta de Hogares por Muestreo
EHPM	Costa Rica's national household survey, the Encuesta de Hogares de Propósitos Múltiples
EHPM	Panama's national household survey, the Encuesta de Hogares de Propósitos Múltiples
EMNV	Nicaragua's national household survey, the Encuesta Nacional de Hogares sobre Medicion de Nivel de Vida
ENAHO	Peru's national household survey, the Encuesta Nacional de Hogares
EHPM	El Salvador's national household survey the Encuesta de Hogares de Propósitos Múltiples
ENAHO	Costa Rica's national household survey, the Encuesta Nacional de Hogares
ENCFT	Dominican Republic's national household survey, the Encuesta Nacional Continua de Fuerza de Trabajo
ENCOVI	Guatemala's national household survey, the Encuesta Nacional de Condiciones de Vida
ENCOVI	The Encuesta Nacional de Condiciones de Vida for Venezuela
ENEI	Guatemala's national household survey, the Encuesta Nacional de Empleo e Ingresos
ENEMDU	Ecuador's national household survey, the Encuesta Nacional de Empleo, Desempleo y Subempleo
ENFT	The Dominican Republic's national household survey, the Encuesta Nacional de Fuerza de Trabajo
ENIGH	Mexico's national household survey, the Encuesta Nacional de Ingresos y Gastos de los Hogares
EPHPM	Honduras national household survey, the Encuesta Permanente de Hogares de Propósitos Múltiples
EPH	Paraguay's national household survey, the Encuesta Permanente de Hogares
EPHC	Paraguay's national household survey, the Encuesta Permanente de Hogares Continua
EPHC	Argentina's national household survey, the Encuesta Permanente de Hogares Continua
GEIH	Colombia's national household survye, the Gran Encuesta Integrada de Hogares
HHS-OLAS	OLAS/SCL Household Survey Dataset

IDB	Inter-American Development Bank
JMP	WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene and their associated dataset for measuring SDG indicators 6.1.1 and 6.2.1a
LAC	Latin America and the Caribbean
LFS	The Bahama's national household survey, the Labour Force Survey
LFS	Belize's national household survey, the Labour Force Survey, differentiated from the Bahama's LFS in text.
MICS	Multiple Indicator Cluster Surveys
LAPOP	Latin American Public Opinion Project
OLAS	Water and Sanitation Observatory for Latin America and the Caribbean
PNAD	Brazil's national household survey, the Pesquisa Nacional por Amostra de Domicílios
PNADC	Brazil's national household survey, the Pesquisa Nacional por Amostra de Domicílios Contínua
SCL	Social Division of the Inter-American Development Bank
SDG	Sustainable Development Goals
SLC	Suriname's national household survey, the Survey of Living Conditions
UNICEF	United Nations International Children's Emergency Fund

Executive Summary

The development of robust public water and sanitation policies requires dependable, transparent, comparable, and representative data; however, the available datasets for the region are complex with indicators that are often misinterpreted by policy professionals. It is essential that users understand the strengths and limitations of these datasets to facilitate source choice and ensure that their use results in accurate insights.

This document discusses and contrasts water and sanitation access data produced by three sources: the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene Dataset **(JMP)**, the OLAS/SCL Household Survey Dataset compiled by the Water and Sanitation Observatory for Latin America and the Caribbean and the Social Division of the Inter-American Development Bank **(HHS-OLAS)**, and data collected by the AmericasBarometer Survey from the Latin American Public Opinion Project **(AB-LAPOP)**.

The document outlines the goals, strengths, and limitations of each dataset and provides a comparative analysis of key indicators, with the aim of giving data users a better understanding of how to use these datasets in tandem to produce a more comprehensive picture of water and sanitation access in LAC.

The quality of many international datasets is depends on the ability of national data ecosystems to generate reliable and methodologically robust data. Many international datasets used in international development, especially in water and sanitation, rely on national figures reported by countries or on microdata generated by countries. The use of national data to produce international datasets has several advantages. First, and most obviously, the task of collecting decentralized information on a global scale is expensive and impractical, especially since data must be updated frequently. Secondly, national statistics agencies generally have the capacity and funding to carry out large-scale country-wide surveys that produce high quality results. These surveys often collect information on relevant topics, so leveraging local data production both reduces cost and decreases the chances of discrepancies and disputes over the accuracy of the numbers.

Simultaneously, national regulators are often required to track data on topics that are difficult to collect otherwise, such as drinking water quality and wastewater treatment data. In many cases these entities are the only centralized source of information on these topics, so having robust data systems within these national entities is key for obtaining decision-relevant information.

Relying heavily on national data for generating international datasets, however, means that countries must collect reasonably similar information. This is a massive challenge as there is a lot of heterogeneity between countries, both in terms of their capacity to collect and manage robust information and in terms of the specific data points they collect. This results in comparability and compatibility challenges in international datasets that many international organizations and regional initiatives work hard to minimize via harmonization methodologies. At the same time, lack of capacity in national data generation creates large gaps both nationally and internationally, that must be filled through supporting and improving the development of robust national ecosystems.

This document compares regional water and sanitation data sources with special attention given to the aforementioned challenges of harmonization and data collection. The goal is to facilitate the use of the available datasets by communicating the major challenges these datasets face and the methodologies used to overcome these challenges.

1. Introduction

The development of robust public policies in the Water and Sanitation sector requires having access to dependable, transparent, comparable, and representative data; however, this simple goal presents many complex challenges. International datasets can be generated via harmonization of national data or generated by independent centralized data generation efforts, both of which have disadvantages. Data generation by centralized regional or international entities can lack the capacity to collect data on the scale that national entities can, resulting data limitations[1]; however, integrating data from national sources into a comparable regional dataset is challenging due to the methodological heterogeneity between countries. To facilitate informed use of available information, it is important to understand these challenges as they apply to existing datasets.

Data describing water and sanitation access in Latin America and the Caribbean (LAC) is produced by international monitoring bodies, international organizations, national statistics agencies, and academic research groups, with each generating data to pursue different goals.[2] These goals are reflected in the methodologies employed, including data collection methods, the frequency with which they collect information, how they define indicators, and how they treat missing values, resulting in distinct strengths and weaknesses in the consequent datasets. Because these datasets have different purposes and strengths, they lend themselves in different uses, and can serve to complement each other with respect to their individual weaknesses, especially related to representativity, consistency, and reliability.

This document discusses and contrasts water and sanitation access data produced by three sources: the Households dataset produced by the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene **(JMP)**, the National Household Survey dataset compiled by the Water and Sanitation Observatory for Latin America and the Caribbean and the Social Division of the Inter-American Development Bank **(HHS-OLAS)**, and data collected by the AmericasBarometer Survey from the Latin American Public Opinion Project **(AB-LAPOP)**. It discusses the goals of the three datasets, their methodologies, and how their objectives relate to their strengths and weaknesses. Descriptive demographic statistics for the two datasets with microdata are compared (HHS-OLAS and AB-LAPOP), to show how sampling within these two surveys may impact the resulting WASH indicators generated from them. Finally, the document presents a comparative analysis of the three datasets based on three key indicators for water and sanitation access:

- Drinking water: Piped improved water network
- Sanitation: Sewer connections
- Sanitation: Septic tank connections

These indicators were chosen due to their presence in all three datasets and their unambiguous definitions relative to other concepts of water and sanitation access, which allows for more direct comparison with fewer caveats across the three datasets.

The analysis reveals that there may be some slight differences between the populations sampled in AB-LAPOP and the national government-run surveys used to produce the HHS-OLAS and JMP indicators, likely related to financial constraints and the resulting smaller sample sizes. The indicator comparison reveals that these differences in the microdata do not have a large impact on the final indicator values in most cases.

1 Temporal or geographic variation of coverage as is the case with UNICEF's Multiple Indicator Cluster Surveys (MICS), or small sample sizes prone to larger margins of error as is the case with the Latin American Public Opinion Project (AB-LAPOP).

2 See Section 2: Characteristics of Water and Sanitation Datasets

2. Characteristics of International Water and Sanitation Datasets

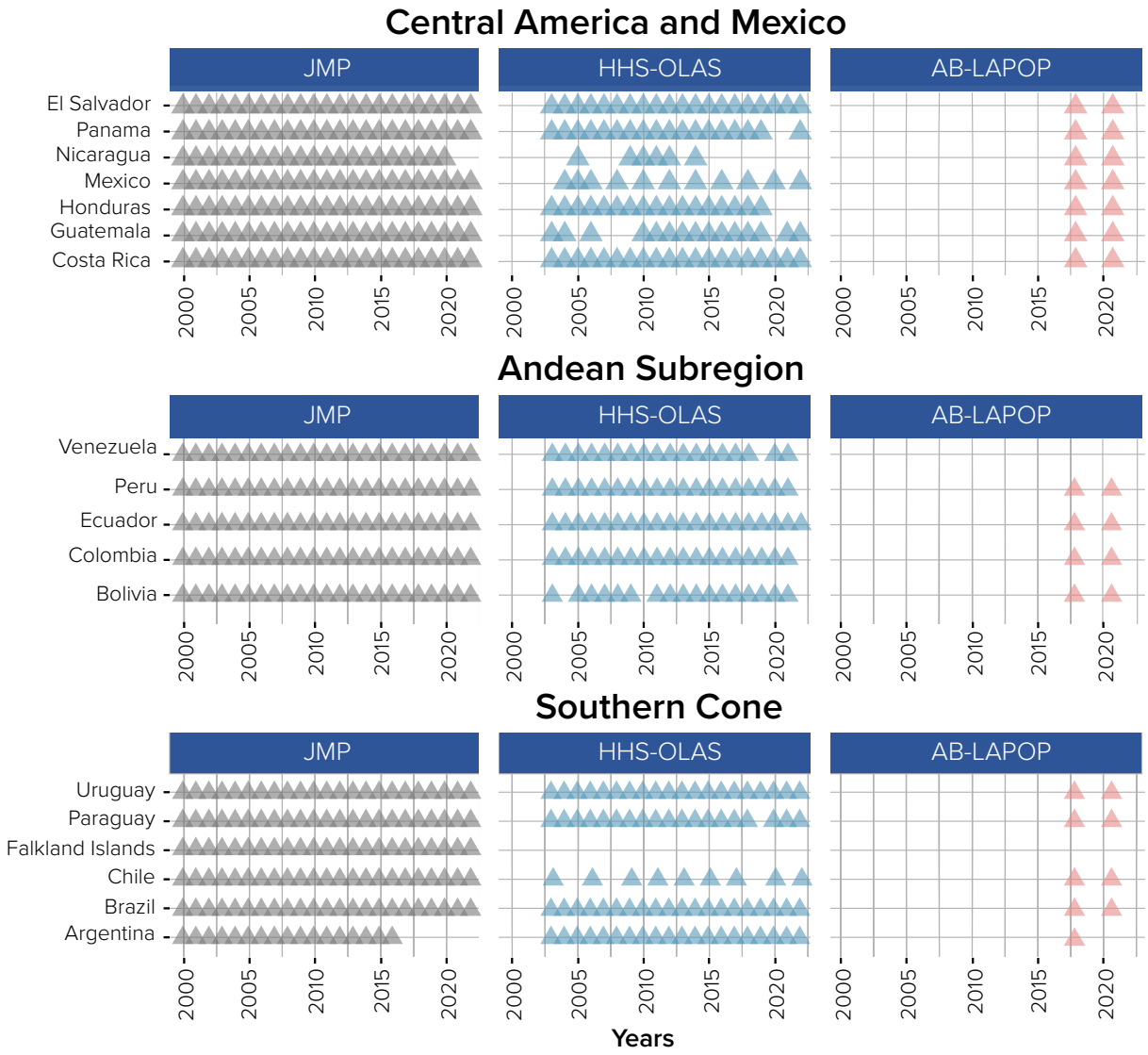
To understand the strengths and limitations of a dataset, it is important to understand the original data sources, the methodology used to generate the dataset, and the dataset's objective. This section discusses these aspects as they relate to the JMP, HHS-OLAS and AB-LAPOP datasets. A summary table of the characteristics, strengths, and limitations of each dataset is in Table 1, while the countries and temporal coverage of the datasets is shown in Figure 1.

Table 1: Summary of datasets and their usage

	JMP	HHS-OLAS	AB-LAPOP
Dataset Location	**Dataset:** • JMP (washdata.org) • Country-level documentation, including country files and inequalities files • Regional-level estimates	**Dataset:** • HHS-OLAS Country-level dataset • HHS-OLAS Regional estimates[3]	**Dataset:** • AB-LAPOP 2018/19 • AB-LAPOP 2020/21 **Microdata documentation:** • LAPOP | Vanderbilt University
General Characteristics	• 22 years of coverage • 49 countries/territories	• 20 years of coverage depending on household survey availability • 26 countries	• Two years of coverage • 20 countries
Advantages	• Complete temporal coverage with linear regression • Regional geographic coverage with estimates for all LAC countries (where data is available) • Global geographic coverage (where data is available) • Official data for measuring SDG 6 progress. • Quantification of missing information • Use of surveys with large sample sizes • Estimates for many Caribbean countries and territories that are not included in other data sources.	• Wide temporal coverage • Regional geographic coverage (where data is available) • Regional estimates • Information on usage of specific water source and sanitation facilities • Seven breakdown dimensions for analysis • Accessible indicators quantification of missing information • Surveys with large sample sizes	• Regional geographic coverage • Comparable without need for harmonization methodologies. • Expansive water and sanitation module.
Limitations	• Only measures indicators and input metrics relevant to SDG JMP framework • Limited comparability due to different microdata methodologies, requiring complex harmonization strategies, see section 2.1.2.2. Comparability • Methodology obscures data gaps, see section 2.1.2.1. Ambiguity Handling	• Limited comparability for select indicators due to different microdata methodologies, requiring complex harmonization strategies, see section 2.2.2.2. Comparability • Gaps in years with no surveys • Low Caribbean coverage and lack of coverage for non-IDB member states.	• Lack of temporal information • Small sample size resulting in larger margins of error. • Note: Possible issues with representativity for very low income, especially in 2018 wave.

3 The countries included in each indicator's estimation is available in the dataset's associated documents.

	JMP	HHS-OLAS	AB-LAPOP
Best for:	• Estimates on progress towards SDG indicators 6.1.1 and 6.2.1a • Focus on services and populations use of services. • Temporal analysis • Regional estimates • Data on concepts not available from household surveys (water quality, wastewater treatment). • Estimates disaggregated by sociodemographic dimensions such as wealth and ethnicity.	• Estimates for rates of access not specific to the JMP framework (specific water sources, sanitation facilities etc.). • Focus on household amenities with distinction between in-home access and public. • Temporal analysis • Regional estimates • Estimates disaggregated by sociodemographic dimensions such as age, sex, income, ethnicity, migration status etc. • Understanding data gaps	• Direct cross-country comparisons • Data on topics rarely covered in national surveys (meter usage, continuity of water access, cost, home treatment).



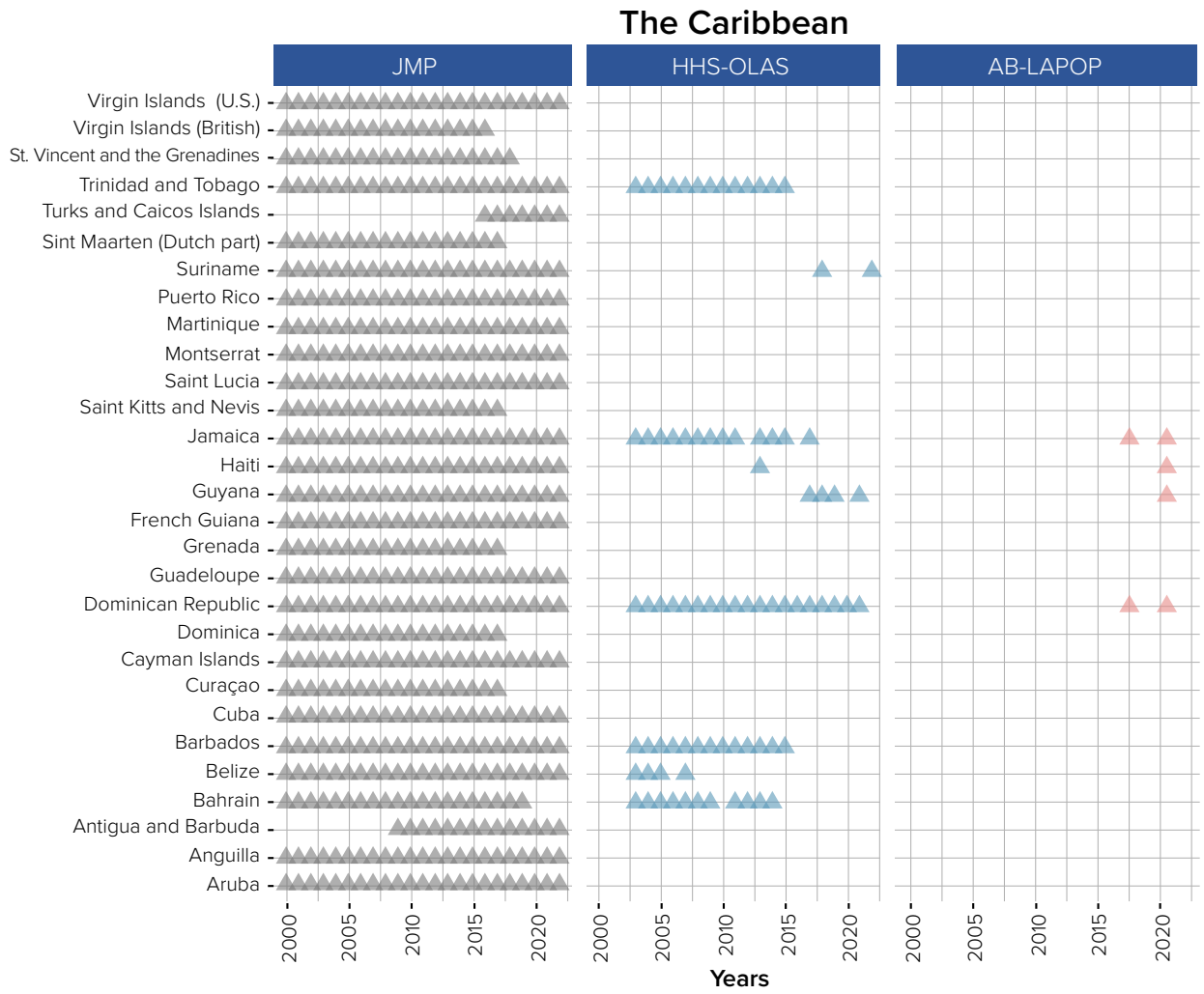


Figure 1: Temporal data coverage by country and subregion, giving data users an understanding of the temporal and geospatial availability of data for each dataset.

The datasets vary in terms of the temporal and geospatial coverage. The JMP has the most complete coverage due to their mandate to provide data for all countries and territories and their methodology which includes imputation for years where data is not available from countries. The HHS-OLAS includes data for the 26 LAC countries that are borrowing members of the IDB[4], excluding non-member states[5] and territories[6]. The AB-LAPOP collects data from 21 countries and has water and sanitation information for two survey waves, one that took place in 2018 and 2019 (2018/19) and one that took place in 2020 and 2021 (2020/21).

Each dataset focuses on specific aspects of water and sanitation services and access depending on their goals. The JMP takes a service perspective, documenting the water and sanitation services that populations use to generate population estimates of access to various levels of water and sanitation services. The HHS-OLAS documents information at the household level and takes an amenity approach, looking at what facilities are available in the home, how they are used and household preferences. The AB-LAPOP also takes a household approach but updates their questions each year and includes general amenity information as well as information specific to certain areas of research such as water and gender roles, household perceptions and opinions of service, and costs. The following sections detail these datasets' goals, methodologies, and indicators so that water and sanitation researchers can better understand their utility.

4 Argentina, Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, the Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Peru, Panama, Paraguay, Suriname, Trinidad and Tobago, Uruguay and Venezuela.

5 Cuba, Saint Lucia, Grenada, St. Vincent & Grenadines, Antigua y Barbuda, Dominica, Saint Kitts and Nevis.

6 Puerto Rico, Guadeloupe, Martinique, French Guiana, Curaçao, Aruba, U.S. Virgin Islands, British Virgin Islands, Cayman Islands, Turks and Caicos, Sint Maarten, Caribbean Netherlands, Anguilla, Montserrat, and Falkland Islands.

2.1 WHO/UNICEF Joint Monitoring Programme Water and Sanitation Data

The WHO/UNICEF Joint Monitoring Programme (JMP) is the custodial agency for monitoring progress towards universal access to safe water and sanitation under the UN framework of the Sustainable Development Goals (SDGs) (JMP, 2023d). Consequentially, one of the JMP's primary tasks is the collection and processing of data on use and quality of water and sanitation services at a global level. The indicators in its dataset are designed to be consistent with the JMP framework for measuring access to water and sanitation and to lend themselves to straight-forward interpretation by decision-makers who may not be familiar with the sector.

Measuring water and sanitation access is a complicated endeavor. Water access must consider physical water access, the quality of the water source (does it comply with drinking water standards?), the time investment needed to procure the water (how far away is the source?), continuity of availability (is it available when needed?), and cost (is it affordable?). To deal with some of these dimensions of access the JMP has defined the measurement of water and sanitation access along a spectrum that runs from a complete lack of access (surface water or open defecation) to the SDG indicators of "safely managed" services. Specific service levels have been defined along this spectrum, conceptualized as a ladder, for both water access and sanitation services (Figure 2) (JMP, 2021a).

Drinking water ladder		Sanitation ladder	
	SAFELY MANAGED Drinking water from an improved water source that is accessible on premises, available when needed and free from faecal and priority chemical contamination		**SAFELY MANAGED** Use of improved facilities that are not shared with other households and where excreta are safely disposed of in-situ or removed and treated offsite
	BASIC Drinking water from an improved source, provided collection time is not more than 30 minutes for a roundtrip including queuing		**BASIC** Use of improved facilities which are not shared with other households
	LIMITED Drinking water from an improved source for which collection time exceeds 30 minutes for a roundtrip including queuing		**LIMITED** Use of improved facilities shared between two or more households
	UNIMPROVED Drinking water from an unprotected dug well or unprotected spring		**UNIMPROVED** Use of pit latrines without a slab or platform, hanging latrines or bucket latrines
	SURFACE WATER Drinking water directly form a river, dam, lake, pond, stream, canal or irrigation canal		**OPEN DEFECATION** Disposal of human faeces in fields, forests, bushes, open bodies of water, beaches and other open spaces or with solid waste

Figure 2: Drinking water and Sanitation ladder defined by the JMP to measure progress on SDGs 6.1 and 6.2

The JMP's 2023 report used data from 1,287 national data sources from 50 countries and territories including all 33 UN Member States in the SDG region of Latin America and the Caribbean. These include 479 surveys with microdata, 192 censuses, 375 administrative data sources and 108 other sources (Table 2).

Table 2: Data sources used to produce the JMP estimates included in their 2023 report, (JMP elaboration with data from JMP, 2023b)

Country or Area	Survey with microdata	Survey without microdata	Census	Administrative data	Other	Total
Anguilla	0	2	2	1	1	6
Antigua and Barbuda	0	3	3	0	1	7
Argentina	10	0	3	1	16	30
Aruba	0	0	5	1	1	7
Bahamas	1	2	2	0	1	6
Barbados	2	1	3	0	1	7
Belize	7	1	3	0	8	19
Bolivia (Plurinational State of)	27	3	3	6	1	40
Bonaire, San Eustaquio and Saba	0	1	1	1	0	3
Brazil	45	0	3	30	0	78
Chile	3	10	3	42	1	59
Colombia	37	1	3	18	1	60
Costa Rica	16	10	3	39	9	77
Cuba	4	1	2	32	1	40
Curaçao	0	0	2	1	0	3
Dominica	0	1	2	0	1	4
Ecuador	18	6	3	5	1	33
El Salvador	28	5	2	14	1	50
Granada	0	2	2	7	1	12
Guadalupe	0	1	15	17	1	34
Guatemala	27	0	3	6	1	37
French Guyana	0	1	14	11	1	27
Guyana	8	0	2	0	1	11
Haiti	8	0	2	1	1	12
Honduras	33	8	3	10	2	56
Cayman Islands	0	1	4	12	0	17
Falkland Islands	0	0	0	1	0	1
Turks and Caicos	1	2	2	1	1	7
British Virgin Islands	0	1	3	0	1	5
U.S. Virgin Islands	0	0	4	4	12	20
Jamaica	17	9	3	2	0	31
Martinique	0	1	16	21	0	38
Montserrat	0	2	3	1	1	7
Mexico	44	1	6	22	1	74
Nicaragua	4	20	1	4	4	33
Panama	9	3	3	8	3	26
Paraguay	18	10	3	0	1	32
Peru	54	1	3	21	4	83
Puerto Rico	11	1	1	5	7	25
Dominican Republic	17	6	3	18	1	45
Saint Kitts and Nevis	0	1	2	1	1	5
Saint Barthélemy	0	0	13	1	0	14

Country or Area	Survey with microdata	Survey without microdata	Census	Administrative data	Other	Total
Saint Martin (Dutch part)	0	0	2	0	0	2
Saint Martin (French part)	0	0	13	1	0	14
St. Vincent and the Grenadines	0	1	3	1	0	5
Saint Lucia	3	2	3	1	1	10
Suriname	5	1	2	2	1	11
Trinidad and Tobago	5	3	3	3	2	16
Uruguay	16	0	2	2	13	33
Venezuela (Bolivarian Republic of)	1	8	5	0	1	15
Total	**479**	**133**	**192**	**375**	**108**	**1,287**

2.1.1 Strengths

2.1.1.1 Flexibility

The definition of distinct service levels allows the JMP to measure water and sanitation service improvements in countries with distinct levels of sectoral development as they "progress" through the ladder's rungs. For example, a country may not demonstrate progress on access to "safely managed" drinking water but show great improvements in provision of "basic" drinking water services, which is defined as use of improved water sources within less than 30 minutes round trip. Service coverage is measured as a percentage of the total population, consistent with the definition of SDG Indicator 6.1.1 and 6.2.1a. This complex framework requires that the JMP collect data on several aspects of water and sanitation services included in each level of their framework. As a result, the JMP not only has information on piped water or sewer system connections, but also estimates of concepts that are harder to measure such as water availability, proximity, quality, and wastewater treatment.

Many countries are unable to produce estimates on safely managed water and sanitation services due to the data required by the methodology. The calculation of safely managed water and sanitation services requires information that cannot be easily obtained from surveys such as water quality information, wastewater treatment data, and waste management from decentralized sanitation solutions (Figure 2). Large data gaps in the region with respect to these topics would hinder measuring progress in LAC, but the inclusion of lower rungs with fewer data requirements allows estimates to be generated. Estimates of basic service coverage – which only require knowledge of the water source and its location, or of the type of sanitation facility used and whether the facility is shared – are available for most LAC countries.

2.1.1.2 Global Coverage

The JMPs goal of measuring global progress towards universal safely managed water and sanitation services requires the dataset to have global coverage. While there are data gaps in countries where the data simply does not exist and estimates cannot be made, the relatively comprehensive geographic coverage of the dataset allows for aggregation of information at the regional and global level, as well as for comparison of information between most countries.

The flexibility of the JMP framework allows the JMP to produce access estimates for most countries for at least some levels of access, including in countries where the information required for monitoring safely managed services is missing. As a result, the JMP dataset includes service estimates spanning 20+ years for most countries around the world, making the JMP data crucial for measuring sectoral change both in terms of progress towards physical access and improvements in the quality of water

and sanitation services around the world.[7]

Additionally, the incorporation of administrative data and data sources beyond household survey data allows important concepts such as water quality and wastewater treatment to be measured. The JMP database incorporates data from a large number of household surveys, censuses, administrative data, and other sources (Table 2). The JMP validates with multiple other datasets and has strict criteria for including a specific dataset in their estimates (JMP, 2018).[8] The JMP database is updated every two years and draft estimates are shared with countries for review and comment before finalization.

2.1.2 Limitations

The task of measuring access to water and sanitation at a global level presents many challenges, but the JMP methodology offers a robust approach that is aligned with its institutional mission of tracking progress in the sector. This methodology, however, may not serve as well for all data needs in the sector, highlighting the importance of looking at sectoral issues through the lens of multiple data sources.

The JMP uses data produced by the countries via censuses, household surveys, as well as national-level administrative data to generate the inputs necessary to categorize the population into each level of access (JMP, 2023a). The information produced from these inputs is specific to the JMP framework of access and does not include information of interest that falls outside this framework. In the case of water access, for example, it does not produce disaggregated estimates of water sources, reliance on bottled water, explicit network connections, or detailed information on differentiation of water sources based on purpose (water used for human consumption vs water used for household tasks). However, the JMP country files and inequalities files do provide disaggregated data from individual national data sources with respect to some water source categorizations. The case of disaggregated data on sanitation access is similar: exclusivity of access regardless of facility type, and explicit use of open defecation is not produced (JMP, 2021a), though again, similar disaggregations are highlighted for individual data sources in the JMP country files and inequalities files.[9] This is largely due to the fact that the JMP framework is designed to take aggregated inputs -- an approach that allows the JMP to more easily harmonize the heterogeneous data available across countries.

2.1.2.1 Ambiguity Handling

Harmonizing information from all over the world presents many challenges, as countries collect sectoral data using different definitions, categorizations, and methodologies. This heterogeneity results in data gaps and the need to create exception handling rules so that indicators can be easily understood. For example, take the case of improved sanitation facilities.

For populations to be classified as using “limited”, “basic” or “safely managed” sanitation services, it is required that they use improved sanitation facilities. The JMP defines “improved sanitation facilities” as “those designed to hygienically separate excreta from human contact” (JMP, 2023c). Improved facilities include some types of latrines such as composting toilets and latrines with concrete slabs, while they exclude other types of latrines such as hanging latrines and pit latrines without slabs (JMP, 2021c).

7 The JMP dataset for monitoring SDG 6 contains estimates going back to 2000. The JMP has been collecting water and sanitation data since the 1990s and was the custodial organization for Millenium Development Goal 7 prior to being charged with monitoring SDG 6

8 The JMP country files include data that are representative of national populations, as well as data that are representative of urban and rural populations. To be considered nationally representative, data should be collected from regions representing at least 80% of the population of interest.

9 The JMP methodology categorizes those without sanitation facilities under “no services/open defecation”. If a household does not have onsite sanitation facilities but does report using offsite facilities it is excluded from this category. Understanding these offsite alternatives (such as reliance on public facilities) may be of interest to policymakers.

In a survey where all latrines are categorized in one response option, e.g. “latrine”, it is impossible to differentiate between the population using improved and unimproved latrines. To deal with this uncertainty, the JMP supplements with other available datasets that are not ambiguous or, if no such dataset is available, the JMP categorizes 50% of respondents using latrines as having “improved” facilities and 50% as having “unimproved” facilities (JMP, 2018). These kinds of assumptions are necessary for the harmonization of different data sources, but they also obscure the uncertainty behind the data, which can be uncovered by considering additional data sources.

2.1.2.2 Comparability

Additionally, harmonization presents certain challenges when comparing data across countries. Differences in response options can alter respondents’ answers. For example, if a household primarily uses bottled water but the national survey does not include bottled water as a response option, it is likely that household will choose their second source, likely piped water from the network. This is not an issue from the perspective of measuring improved water access as both are improved sources, however, it will result in that country having a higher rate of piped water use than a similar country where the bottled water option was offered. The JMP recommends that when households report using bottled water as their main source of drinking water, that households should be asked about their main source of water for other purposes. When this is done (for example, in most MICS and DHS surveys), the JMP adjusts the estimates of piped water coverage to include households that use bottled water but have piped water coverage. However, this secondary question is not included in all household surveys.

This also comes into play with respect to the definition of variables. The JMP uses the questions available in each survey, but these questions can be measuring different concepts. For example, household surveys in Brazil collect data on piped water from any source (public network, well, spring, delivered cistern), while in the Dominican Republic, household survey collect information specifically about piped water from the public network.[10] The JMP classifies both as ‘piped water’. Examples of this variation can be seen throughout the dataset, especially with respect to concepts that are difficult to measure, such as continuity of access.[11]

Although these differences do not preclude comparison between the countries, it is essential that users understand how to use the documentation to be able to interpret comparisons between countries accurately.

2.1.3 Key Takeaways

The goal of the JMPs dataset is to provide standardized information that can be used to measure progress towards SDG 6.1 and 6.2. The indicators within the dataset have good temporal and geographic coverage and allow for country level, regional, and global estimates that are comparable over time. As the official Monitoring Body for SDG 6.1 and 6.2, the dataset is widely used and is accepted as a quality source of information.

The JMP collects data specifically to measure water, sanitation, and hygiene access globally according to its official framework. As a result, the dataset focuses exclusively on topics directly related to that framework and the eventual generation of access estimates within the framework of the JMP access ladders. The charge of producing estimations on a global scale requires certain simplifications to deal with wide-scale harmonization, the implementation of imputation techniques and the use of various data sources in the case of missing information.

10 “Del acueducto”

11 Limitations on comparability are discussed in detail in the HHS-OLAS section but apply to both the JMP and HHS-OLAS. Additionally, the availability of water as needed is a complex topic. The JMP collects data on “availability when needed” which is related to continuity, defining it as “availability as needed if explicitly stated” or available more than half the time. It’s possible for households to have water “available when needed” even if piped water is not continuous if there is adequate storage within the household.

As a result, the JMP dataset provides high-quality indicators from nationally produced data for use when the following is needed:

- Estimates of progress towards SDG indicators 6.1.1 and 6.2.1a
- Estimates of access at high levels of geographic aggregation, e.g., regional access
- Estimates of country, regional or global-level access over time
- Estimates measuring elements of access specific to the JMP framework (e.g., Water source accessible on premises, available when needed, free of contamination)
- Estimates at subnational administrative levels (subnational regions and/or states) urban/rural and income levels.[12]

The JMP dataset is less appropriate when looking at:

- Topics outside of the scope of the JMP framework (e.g., network connections, water sources for water differentiated by use case, water metering)
- Subnational information on access related to ethnicity, disability, or migratory status

2.2 OLAS/SCL Household Survey Dataset

The OLAS/SCL Household Survey Dataset (HHS-OLAS) was produced by the OLAS in collaboration with the Social Division of the Inter-American Development Bank using microdata from National Household Surveys throughout LAC. The Social Division harmonizes this microdata according to definitions established by the OLAS and then uses it to generate indicators on water and sanitation access. The goal of the dataset is to close data gaps related to water and sanitation by providing detailed information on types of water sources and sanitation facilities used in the region, as well as various sociodemographic dimensions that can be used for data analysis. The dataset also attempts to expose data gaps by measuring uncertainties in the data collected by the national household surveys. The dataset is meant to serve as a complementary source to the information produced by the JMP. The dataset integrates household survey data from 23 countries throughout the region with information spanning from 2003 to 2022, with an average of more than 13 years of data per country.

A full list of the surveys integrated in the current dataset is below (Table 3). Indicators are calculated in household percentages and total number of households while population-level information could be extrapolated but is not available as part of the dataset. The data is updated multiple times per year as new harmonized surveys are produced by the Social Division of the IDB.

12 Data available in JMP inequalities files.

Table 3: Data sources for the National Household Survey dataset (HHS-OLAS) by year at time of writing.

Country	Survey	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
Argentina	EPHC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	20
Bahamas	LFS	■	■	■	■	■	■	■		■		■	■									10
Belize	LFS		■	■		■																3
Bolivia	ECH	■		■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■		17
Brazil	PNAD	■	■	■	■	■	■	■		■	■	■	■	■								12
Brazil	PNADC														■	■	■	■				4
Chile	CASEN	■			■			■		■		■		■		■			■			8
Colombia	GEIH				■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		16
Costa Rica	EHPM	■	■	■	■	■	■	■														7
Costa Rica	ENAHO								■	■	■	■	■	■	■	■	■	■	■	■	■	13
Dominican Republic	ENFT	■	■	■	■	■	■	■	■	■	■	■	■	■	■							14
Dominican Republic	ENCFT															■	■	■	■	■		5
Ecuador	ENEMDU	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	19
Guatemala	ENEI	■	■						■	■	■	■	■	■	■	■	■	■		■	■	14
Guatemala	ENCOVI				■																	1
Honduras	EPHPM	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■				17
Haiti	ECVMAS										■											1
Mexico	ENIGH		■	■	■		■		■		■		■		■		■		■			10
Nicaragua	EMNV			■				■					■									3
Panama	EHPM												■	■	■	■	■	■			■	7
Peru	ENAHO	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		19
Paraguay	EPH	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■						15
Paraguay	EPHC																■		■	■	■	4
El Salvador	EHPM	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	20
Suriname	SLC															■						1
Trinidad and Tobago	CSSP	■	■	■	■	■	■	■	■	■	■	■	■	■								13
Uruguay	ECH	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	19
Venezuela	EHM	■	■	■	■	■	■	■	■	■	■	■	■	■								13
Venezuela	ENCOVI														■	■	■		■	■		5
	Total	**16**	**15**	**17**	**18**	**16**	**16**	**17**	**14**	**17**	**17**	**17**	**19**	**17**	**16**	**17**	**16**	**13**	**12**	**12**	**8**	**310**

■ Incorporated

To generate the dataset, harmonized variables are created within the microdata for each national survey, then used to generate indicators incorporating the weights and/or expansion factors as specified by the national survey documentation. The water and sanitation indicators follow many JMP definitions, such as those of improved water sources and sanitation facilities and measures elements of access included in the JMP concepts (such as continuity of water availability as a subcomponent of water access and exclusive use of sanitation facilities as a subcomponent of sanitation access).

The dataset differs from the JMP methodology in some key respects, including how certain indicators are measured and how exceptions are handled. These differences will be discussed below.

2.2.1 Strengths

2.2.1.1 Granularity of information

The HHS-OLAS dataset follows the JMP framework for defining access, using, for example, the same definition of "improved" to categorize water sources and sanitation facilities. This allows for easy comparison with the JMP dataset and facilitates usability; however, due to its more general sectoral goals, the HHS-OLAS dataset contains many indicators beyond those included in the JMP dataset. The water source and sanitation facility indicators look at access through the lens of household amenities (availability in-home is distinguished from amenities accessed in a public setting), as opposed to the JMP's indicators' organization around services used. The indicators include indicators of specific relevance in LAC, such as reliance on bottled water or trucked water, and indicators examining water source by use case (drinking water or human consumption vs household water use). This region-specific focus and broader thematic coverage make the HHS-OLAS an important complimentary source of information. Additional topics covered by the HHS-OLAS dataset include but are not limited to the following:

- Granular information about water sources (both for general use and human consumption):
  - » Public standpipes
  - » Bottled water
  - » Rainwater harvesting
  - » Trucked water
  - » Surface water
- Additional information about water availability:
  - » Percent of homes that report no water shutoffs
  - » Homes that report having sufficiently consistent access to water
  - » A combination of the above that fits the JMP definition
- Treatment (perception of water quality):
  - » Whether a household treats their water before consuming it
- Metering:
  - » If a household has a water meter
- Lack of sanitation access:
  - » What households do if they do not have access to sanitation (use public or neighbors' toilet, practice open defecation)
- Information not addressed in the countries' surveys:
  - » Does not ask about water for human consumption
  - » Does not address continuity of water availability
  - » Does not ask about exclusivity of sanitation facilities

The harmonization process uses consistent definitions across countries to generate a comparable dataset with respect to these indicators and a variety of dimensions, allowing users to breakdown information at a subnational level. These dimensions currently include zone (urban/rural), and income quintile, gender, age, ethnicity, disability status, and migratory status, allowing for detailed subnational analysis of how sociodemographic factors and water and sanitation access interact.

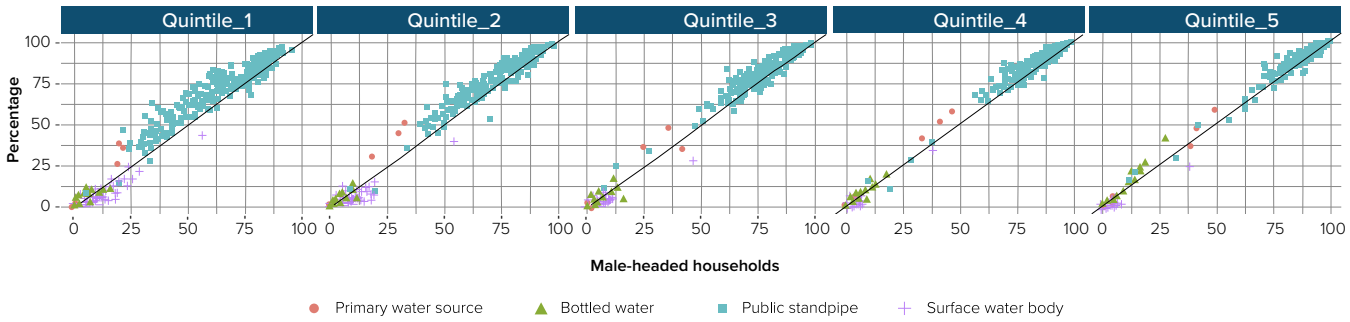


Figure 3: A visualization of rates of access for some water sources in male headed households vs female headed households by income quintile (all years, all countries), demonstrating analyses available with the HHS-OLAS data. Female headed households generally have higher access rates for improved water sources, and lower access rates for unimproved water sources when compared with their male counterparts.

2.2.1.2 Harmonizing response options and comparability

The dataset requires the harmonization of information from over 20 national-level data sources, which results in the need to establish exception handling rules for situations where household surveys collect information on a specific indicator in an unexpected or difficult to classify manner. The dataset is designed to manage these differences by integrating them in indicators that measure uncertainties. For example, there are common cases where the response options provided in a survey cannot be easily classifiable under the JMP definitions of improved water sources. To deal with this uncertainty, there is a specific indicator for the percentage of households with unclassifiable water sources. In other cases, surveys simply do not address a topic, creating a gap in the dataset. For such cases, a variable showing this gap was created so that the gaps can be quantified.

For example, Figures 4 and 5 show general water sources used in Colombia and Bolivia from 2013 to 2020. This information can be compared across countries, but as a rule it should be compared in the full context, that is, looking at all water sources. This is because in Colombia, roughly 5% of households' sources cannot be classified as definitively improved or unimproved, while in Bolivia virtually all water source response options fall clearly in one of these categories. This strategy allows for comparison across countries where data is available. For example, if there are values for the percent of households that use bottled water, that variable can be compared across countries. However, the fields for "Other improved sources", "Other unimproved sources" and "Uncategorizable sources" should be visualized together with other water sources if inter-country comparison is necessary.

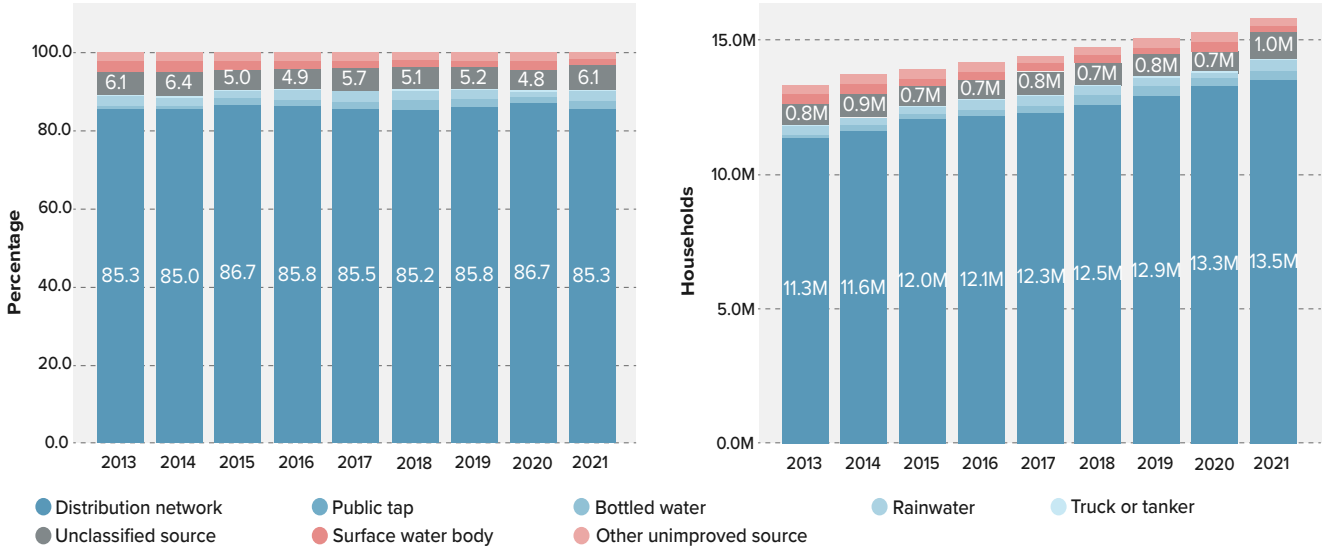


Figure 4: Water source access, Colombia

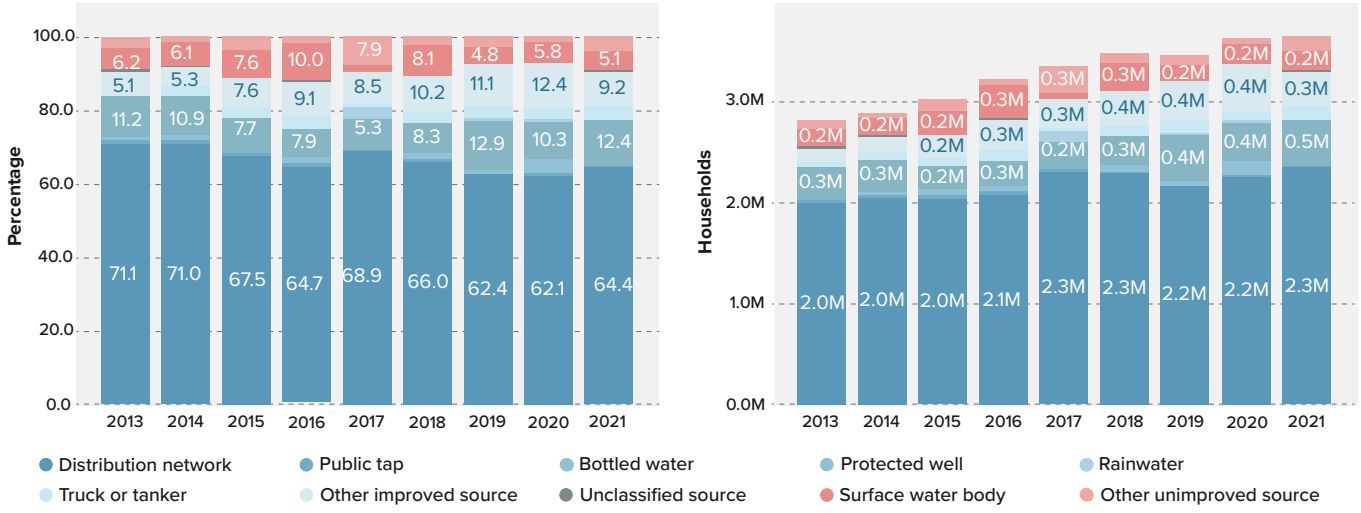


Figure 5: Water source access, Bolivia

When comparing data across countries, it is useful to visualize it in the context of other variables to facilitate understanding. For example, in Figure 6 we can see data from all countries that have general water source data for 2021. We can definitively compare data on distribution network access between all countries as all countries address this topic in their surveys' questions and response options. From looking at the graph, however, one can see that Argentina does not address protected wells as a response option. Those who use protected wells in Argentina are included under "Unclassified source" as the survey's response options include "well" without specifying if it is protected or not. Looking at these variables together provides users with a richer understanding of what the indicators in the dataset represent.

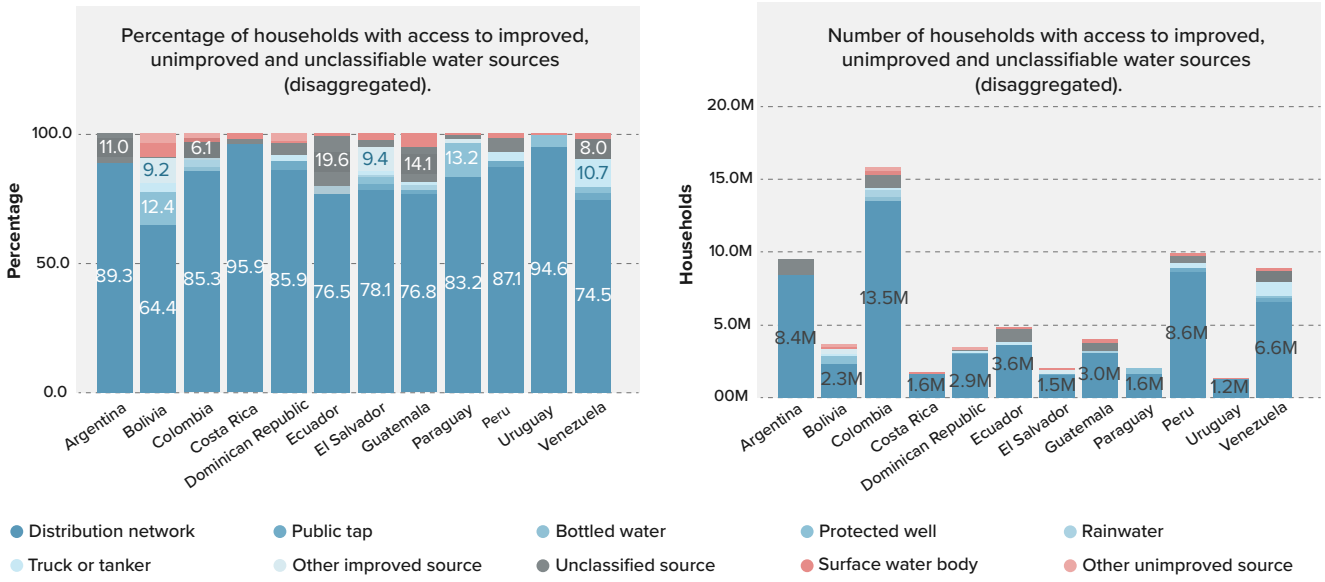


Figure 6: Primary water sources, country comparison (OLAS-HHS dataset, 2021)

2.2.1.3 Highlighting ambiguity and gaps

Viewing information on data sources is especially useful when visualizing information that commonly has data gaps. One such area is the subject of water sources used for human consumption, as there are some countries that only ask this information for certain water sources. For example, Peru only asks households that get their water from the distribution network if the water is used for human consumption, while El Salvador only has this information for users who drink bottled water. By including information on where the known data gaps are, users can see that this gap is conscientious, not a data error.

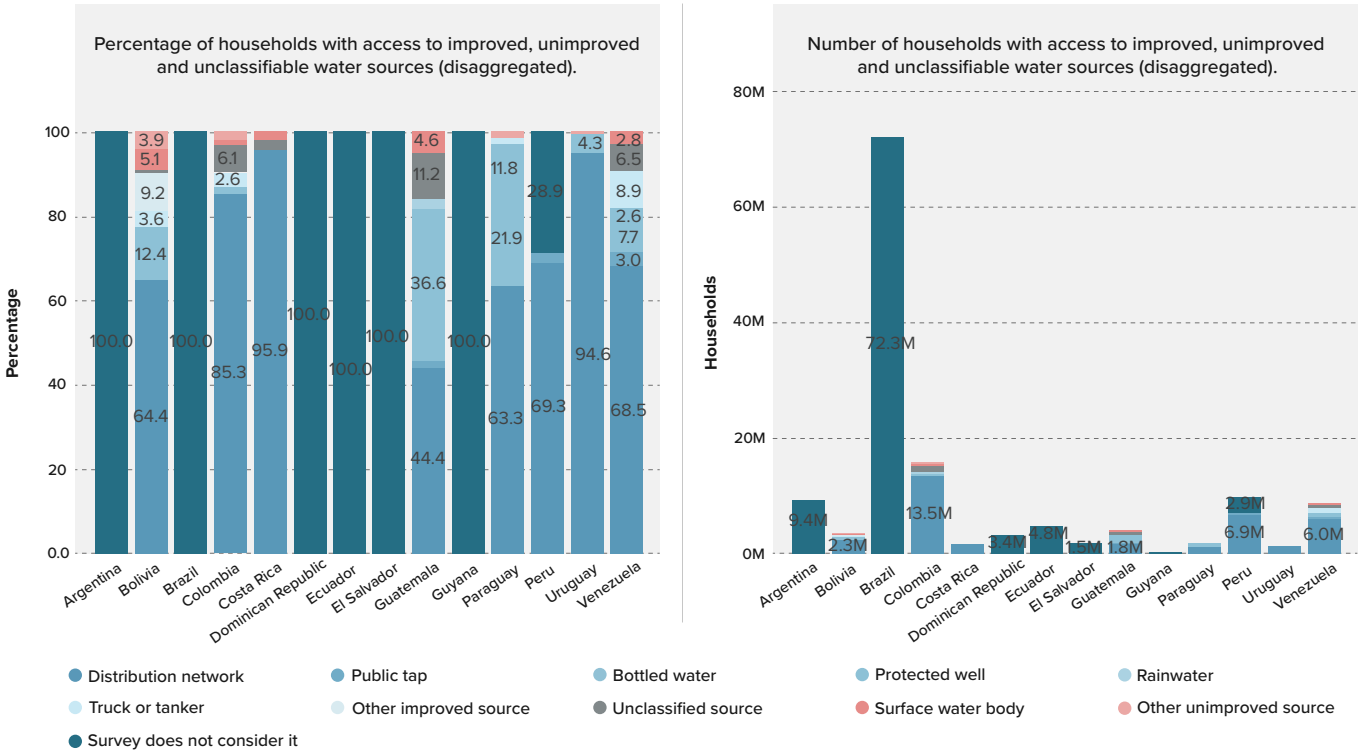


Figure 7: Primary water sources for human consumption, country comparison (OLAS-HHS dataset, 2021)

2.2.2 Limitations

2.2.2.1 Temporal availability

The HHS-OLAS dataset has a wide temporal range, however data for each country is only available for years in which the country conducts a National Household Survey which includes questions on water and sanitation. As a result, cross-country comparisons are not available for all countries every year. This information could be generated from the dataset via a regression methodology like that employed by the JMP, but it is out of the scope of the dataset in its current iteration.

2.2.2.2 Comparability

Since the HHS-OLAS dataset combines information from surveys throughout the region, it faces similar challenges to the JMP when it comes to harmonizing information. Different questions and response options require categorizations that allow for these differences to be incorporated in a way that is consistent but also remains simple enough to be interpreted easily. These differences also produce data gaps which the HHS-OLAS categorizations attempt to manage transparently, as previously discussed. While the dataset's strategies for handling ambiguity increase the comparability between countries there are some variables that should not be compared.

Indicators on **continuity of water access should not be directly compared** because surveys use different time periods and metrics for collecting data on this topic. For example, countries may ask how often in the last week, month, or 3 months the household has had water service or access to water. Sometimes these questions are measuring the consistency of piped network water service, while sometimes they are asking about all water sources.[13] Response options also vary; sometimes answers are given in hours per day, days per week or days per month, making it difficult to directly compare.

13 Specifically, Brazil, Bolivia, Colombia, Mexico, Panama, and Paraguay as about continuity only for those respondents that report having network access, otherwise the question addressing continuity is skipped by the interviewer. El Salvador asks about continuity for respondents that have piped water from any source, while Peru and Venezuela ask all respondents about continuity, regardless of reported water source.

The HHS-OLAS dataset has three variables dealing with this topic, one that measures respondents that report no water cutoffs during the specified period, one that measures that households report sufficient access or have access half of the time or more (consistent with the JMP definition (JMP,2018)), and one that indicates that this question is not asked. For cases where continuity of water availability is only asked to households that used network water as their primary source, all other households are categorized under the indicator that signals that the question was not asked. This allows users to compare the data between countries on this topic within the full context, being able to visualize what is unknown. This information should not be directly compared between countries except to give a general context and even then, it should be viewed holistically (e.g., households that have not reported cutoffs and those that were not asked the question). These indicators can, however, be used for temporal comparisons within countries in most cases.

Variables on **water treatment should also not be compared** between countries but can be used for temporal analysis within countries. This is because countries address the issue of treatment differently, sometimes only asking about certain water sources (piped network water vs all water sources) or certain methods of treatment (filters versus any treatment).

2.2.2.3 Survey changes

Sometimes, due to discontinuation of a survey or changes to a national survey, there are significant changes in how a topic is approached. An extreme example of this can be seen in Venezuela, where, due to the situation in the country the survey used was changed from the Encuesta de Hogares por Muestreo (EHM, 2006-2015), to the Encuesta Nacional de Condiciones de Vida (ENCOVI 2016 to present). The EHM asks only if households filter their water, while the ENCOVI asks about any type of water treatment, resulting in a large increase in the value of the indicator. (Figure 8). The HHS-OLAS dataset has a field that specifies the survey source for each data point, so these changes are transparent, but it is an added level of complexity for users. Users need to be aware of these changes so that they can make the appropriate corrections to their dataset to serve the needs.

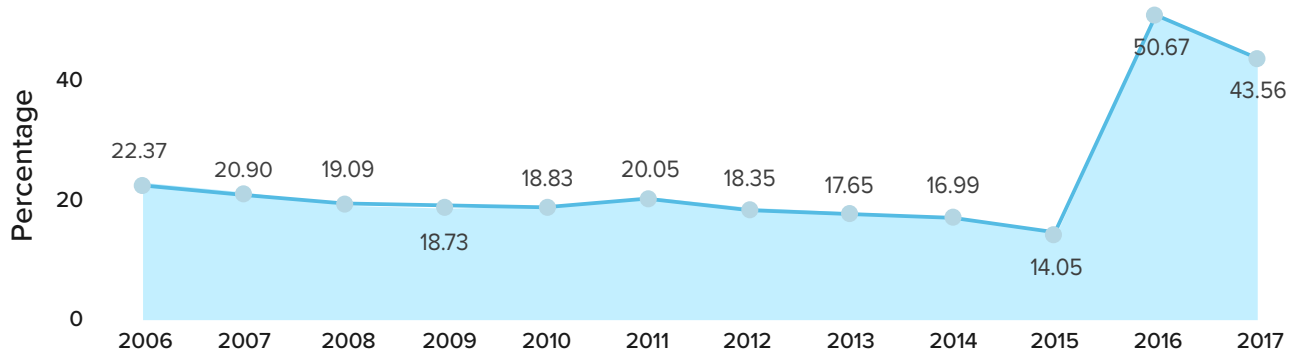


Figure 8: Data on at-home water treatment showing a survey change in Venezuela.

2.2.3 Key Takeaways

The goal of the HHS-OLAS dataset is to provide standardized and transparent water and sanitation access information for LAC that can be used by researchers and policy makers to better understand the sector. The level of detail of the indicators with respect to water sources and sanitation facilities and the dimensions provided for analysis support this goal. As a result, the HHS-OLAS dataset provides high-quality data for use when the following is needed:

- Estimates that are disaggregated by sociodemographic dimensions such as zone, income, ethnicity, or migration status.

- Estimates for indicators not included in the JMP framework, such as household amenity prevalence, use of groundwater, bottled water, and public sanitation facilities in select countries.
- Estimates of access at high levels of geographic aggregation, e.g., regional access
- Temporal analysis
- Comparisons[14] between countries for countries with data from the same year.

The dataset is not well-suited for:

- Making statements related to SDG progress
- Measurement of service use (populations that use specific services)

2.3 Latin American Public Opinion Project's AmericasBarometer

The AmericasBarometer (AB-LAPOP) is the Latin America Public Opinion Project's largest survey research project. It is a periodic survey of 34 countries and includes a common core questionnaire as well as country-specific modules collecting information on public opinion and experiences on topics including governance, education, health, environment and living conditions.

The AB-LAPOP is a comparative survey that covers all independent countries in North, Central and South America, as well as several countries in the Caribbean, with the goal of creating a temporal dataset for the western hemisphere that includes a broad range of topics, produced by a standardized methodology and the use of a common core questionnaire which permit valid comparisons across countries and over time (LAPOP, 2023a). For the 2018/19 wave and the 2021 wave, the Inter-American Development Bank financed the inclusion of questions on water and sanitation access, covering topics that are often not included in national household surveys. This is especially true of the 2021 wave, which included question on water costs and intrahousehold gender dynamics as they relate to water-related tasks.

2.3.1 Strengths

2.3.1.1 Regional Consistency

When compared to the other sources included in this report, the implementation of a consistent methodology (survey questions, response options, survey design) across the region presents a clear advantage. Researchers can work directly with the microdata collected without having to make assumptions or create round-about strategies to generate harmonized indicators from dissimilar data. Relatively consistent categorization of water sources and sanitation facilities, as well as consistent approaches to questions such as continuity of water availability, facilitate comparisons for country-level metrics across the region. Additionally, all information in each survey round is collected during the same period in contrast to the household surveys which are used by both JMP and HHS-OLAS and take place with varying frequencies throughout LAC.

2.3.1.2 Question specificity

The Water and Sanitation Division of the Inter-American Development Bank has worked closely with the AmericasBarometer project to expand the section on living conditions included in the survey and to collect information more aligned with the status of the sector. Because the survey is run via a centralized structure, it is easy to incorporate new questions for all countries throughout the region, allowing the IDB to get information that is not included in most national surveys in LAC. This includes information on water costs, water quality and at-home treatment, and upkeep of onsite sanitation solutions.

14 Excluding the exceptions mentioned in section 2.2.2.2

One important example of this is the fact that the AB-LAPOP differentiates between water sources used for drinking and water sources used for other purposes. This is an area of interest in LAC because many countries, such as Mexico and the Dominican Republic, have very high bottled water consumption rates, despite having high rates of access to the public distribution network. Differentiating between water sources used for drinking versus those used for other purposes is an important distinction which is not available for all countries.[15] The HHS-OLAS dataset includes this differentiation but because the household surveys themselves often do not differentiate there are large data gaps with respect to the comparison of water source by use. The AB-LAPOP, therefore, is useful as a regional temporal cross section for data concerning specific subjects.

2.3.2 Limitations

2.3.2.1 Sampling limitations

The AB-LAPOP surveys have smaller sample sizes than the household surveys (used by HHS-OLAS and JMP) or national censuses (used by JMP) resulting in national estimates being less robust than those generated from other sources. This is supported by differences in the microdata and demographic statistics generated from the AB-LAPOP dataset in comparison with those of the national household surveys and a few significant differences in indicator values (see section 3.1 and 3.2). The analysis indicates possible over-representation of the lower-middle to middle class and an under-representation of the lower and upper classes in 2018. A similar pattern is seen in 2021 but is much less pronounced. This sampling bias does not seem to have a significant impact on most water and sanitation indicators; however, it may affect indicators that do not rise linearly with economic status. This is supported when looking at values for open defecation/lack of access between the datasets. The 2018 AB-LAPOP values are much lower than the JMP and HHS-OLAS values.

Figure 9: The LAPOP 2018 has very low values for open defecation/lack of access when compared to the JMP and HHS-OLAS, while the LAPOP 2021 values seem to correspond more closely.

2.3.2.2 Lack of temporal data

Another limitation is the lack of temporal information on water and sanitation data. The AB-LAPOP's primary goal is collecting information about public opinion throughout the region on a variety of public

15 This distinction is made in some country household surveys such as those of Colombia, Guatemala, and Bolivia, as well as in UNICEF's Multiple Indicator Cluster Surveys (MISC) and the Demographic and Health Surveys from the DHS Program. These surveys, however, are carried out in a few countries in LAC per year making cross country comparisons difficult.

policy subjects. Although the AB-LAPOP began in 2008, detailed water and sanitation data were not incorporated until 2018/19 making it inappropriate for temporal comparison.[16] The fact that only two years of data are available makes it difficult to assess the reliability of the data, as the other datasets can be compared using linear regressions, which serve to "smooth out" any irregular years. This is particularly relevant because the Covid-19 pandemic resulted in methodological changes in the 2021 wave that could have impacted the continuity of the dataset (LAPOP, 2021). Specifically, the need for social distancing resulted in the interviews being carried out exclusively using mobile telephones (Montalvo et al., 2022). Although the sampling strategy was adjusted to account for the move to mobile phone interviews, it is possible that this change has resulted in some discontinuities between the 2018/19 wave and the 2020/21 wave.

2.3.3 Key Takeaways

The AB-LAPOP's centralized management structure produces directly comparable information from countries throughout the region. The survey includes the same questions and response options and conducts the surveys during roughly the same period each wave, allowing for the generation of data that is easy to harmonize and work with. Additionally, its central structure means that it is easy to add questions on specific topics, making it a valuable resource for collecting information on data gaps.
The AB-LAPOP data is best suited for:

- Direct cross-country comparisons, especially of data on continuity of water access
- Research analyzing water costs, perceptions of water quality, and other topics with low data availability at the regional level.

It is not well suited for:

- Temporal analysis of water and sanitation variables
- Making definitive statements about access
- Making statements about SDG progress

16 This also has an impact on the indicator comparison, which should be smoothed out via linear regression but could not be performed because of lack of datapoints.

3. Dataset comparisons

An in-depth comparison of AB-LAPOP and household survey microdata was conducted, as well as a comparison of final indicators across all three datasets. This section details the results of these two comparisons and highlights key takeaways. Differences in the populations sampled and the representativity of the surveys as well as differences in methodology and definitions can result in different values for final indicators. This exploration serves to identify key differences between the three datasets and communicate those to users.

3.1 Microdata Comparison

Differences in microdata due to sampling strategies and survey design can have large impacts on the final calculation of indicator values. This section compares demographic information from the AB-LAPOP and household survey microdata17 from surveys conducted for 2018/19 and 2020/21 using weighted samples. The purpose of this comparison was to explore if the two survey groups were representing sufficiently similar populations.

A basic statistical comparison of the HHS18 and AB-LAPOP microdata shows significant differences between populations. Table 4 shows some key statistics in each set of data, taking survey design into account. On a macro level, the populations surveyed in 2018 seem similar between surveys, with education levels, income, and urban representation being close across surveys. Exceptions to this are the age of the head of household, which is consistently 10 years younger in the AB-LAPOP data, a difference that applies both to 2018 and 2021. 2021 data also show significant differences in income and urban representation, with the AB-LAPOP having lower average income across countries and lower urban populations when compared to the Household Survey microdata. This is important because both the National Household Surveys and the AB-LAPOP strive to be representative at a national, urban, and rural level.

Table 4: Average and standard deviations of country-level calculations. Means were calculated at the country level accounting for survey design and then averaged for each dataset, 2018/19 and 2020/21.

Variable	Year	AB-LAPOP	HHS	Average Difference (HHS minus AB-LAPOP)
Age	2018	40.59	50.31	9.72
	2021	40.00	50.59	10.57
Education Level*	2018	2.89	2.79	-0.10
	2021	1.98	1.79	-0.20
Income USD	2018	509.20	529.24	20.04
	2021	422.61	455.77	33.15
Urban	2018	0.72	0.72	0.01
	2021	0.74	0.77	0.03

**There were significant changes between educational categories in the LAPOP between 2018/19 and 2020/21 (Table 5).*

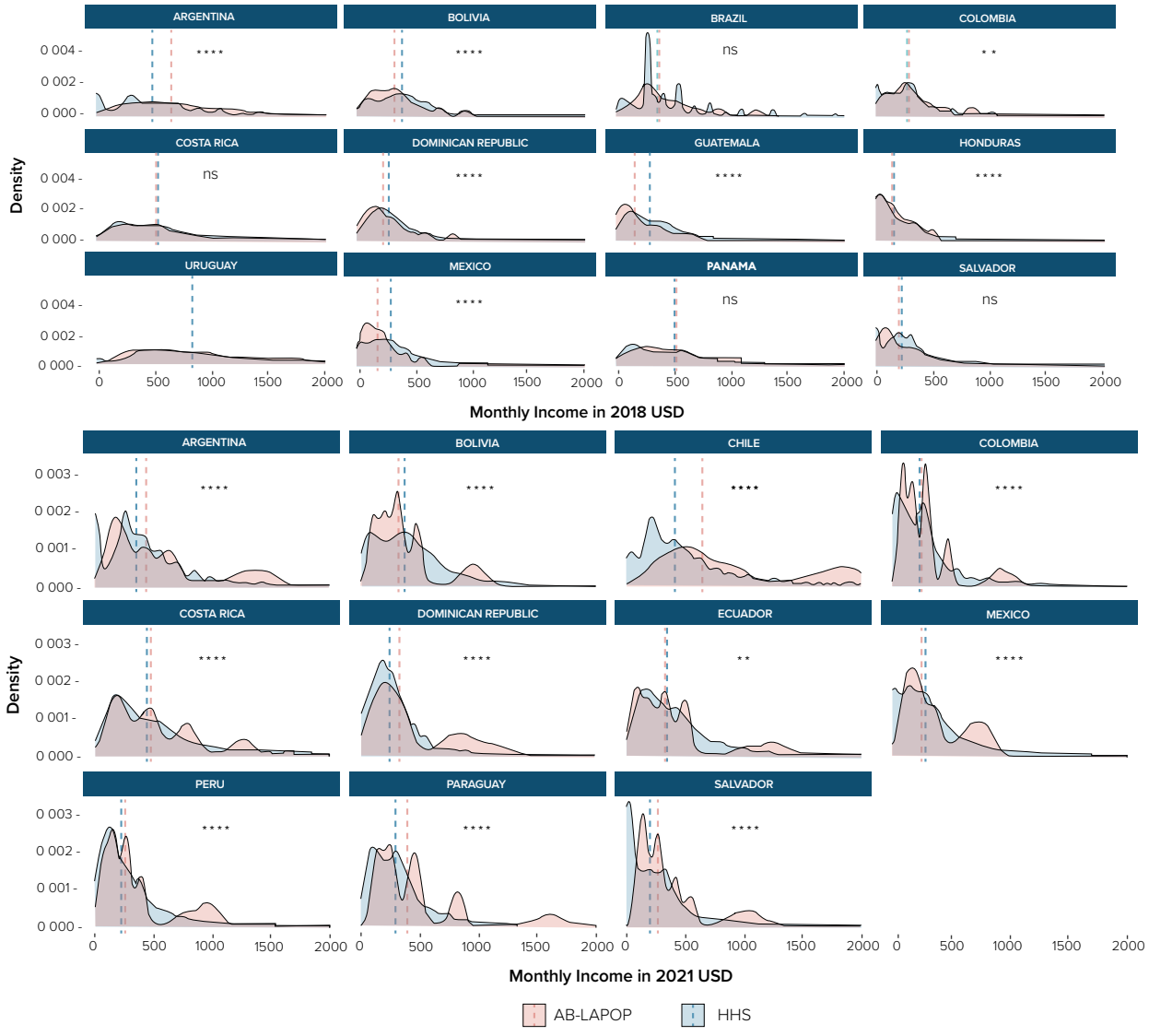
Table 5: Education categories, 2018/19 and 2020/21

Value	Description 2018/19	Description 2020/21
0	No education	No education
1	Some primary school	Primary (incomplete or complete)
2	Completion of primary school	Secondary (incomplete or complete)
3	Some secondary school	Post-secondary (incomplete or complete)
4	Completion of secondary school	
5	Some or completion of post-secondary	

17 National Household Surveys are the key microdata used by both the JMP dataset and the HHS-OLAS.

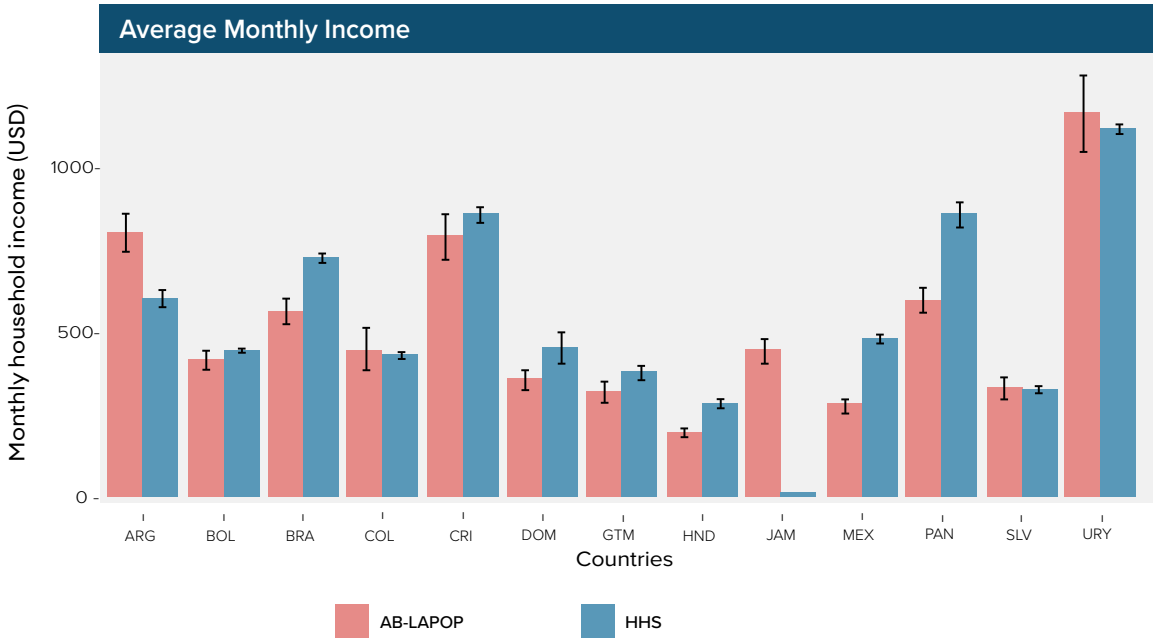
18 HHS refers to the microdata from the national household surveys.

Despite similar averages across country level data, there are significant differences in the distributions of these variables. With respect to income, we see that the distributions of income are significantly different between the two samples for all countries for 2021 and all countries except Brazil, Costa Rica, Panama, and El Salvador for 2018/19. This could be, in part, due to the imputation process applied to the AB-LAPOP data. The AB-LAPOP dataset only provides income data in ranges, so continuous values were imputed (Canavire-Bacarreza et al., 2022). For more information on this process see "Working Paper No IDB-WP-01571: Water Affordability Measures Under Multiple and Non-Exclusive Sources in Latin America and the Caribbean by Roberto Martínez-Espiñeira y María Pérez Urdiales. The 2018/19 round includes sixteen income range categories, while the 2020/21 survey only includes five, reducing the accuracy of the imputation process. The distributions tend to show less low income and less high-income representation in the AB-LAPOP, although this trend is not consistent across all countries. AB-LAPOP data from Guatemala and Mexico notably seem to have more representation of poorer households and less representation of rich households (Figure 10). Country-level average incomes are similar across datasets (Figure 11).

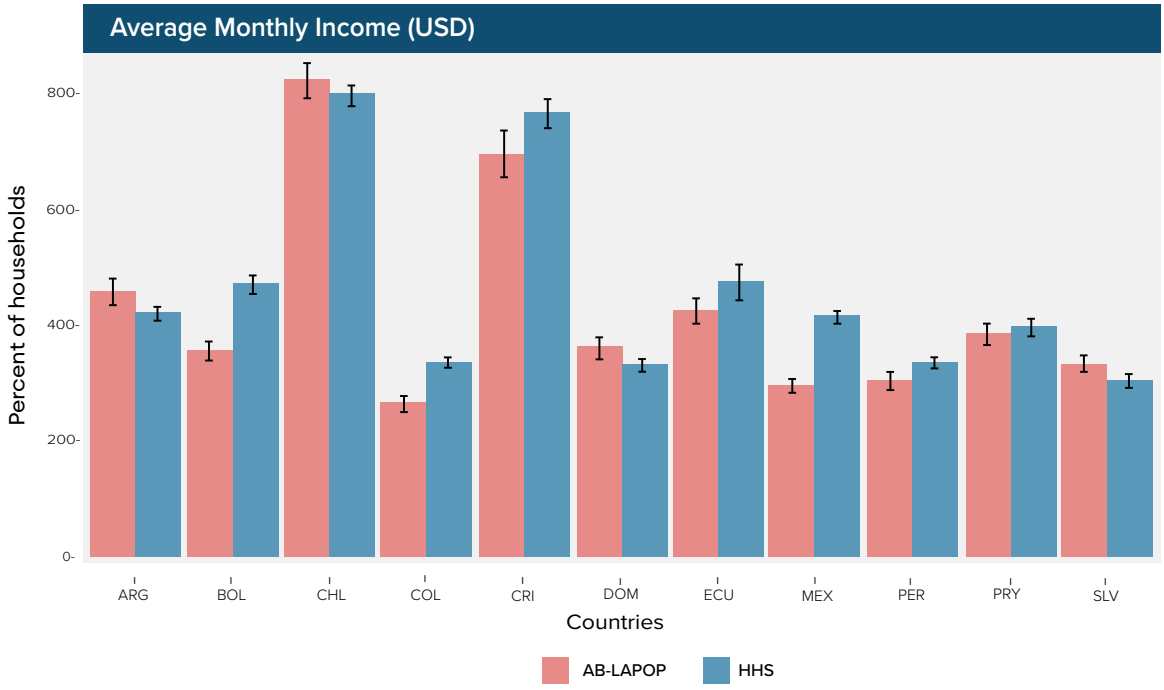


The asterisk in the center of each graph indicates the significance level of a Wilcoxon ranked sum test between the two surveys.
The dashed vertical line represents the median of each dataset.
*Descriptive statistics for Argentina's National Household Survey are only for the urban population.
Source: AmericasBarometer 2021 and National Household Survey 2020.

Figure 10: Income density curves for 2018 and 2021 between AB-LAPOP and the HHS microdata



*Descriptive statistics for Argentina's National Household Survey are for the urban population.
Jamaica does not include Income information in their household survey.
Source: AmericasBarometer 2018-2019 and National Household Surveys 2018-2019.



*Descriptive statistics for Argentina's National Household Survey are for the urban population.
Source: AmericasBarometer 2020-2021 and National Households Surveys 2020.

Figure 11: Average monthly income between the two micro-datasets

The most significant differences between the demographics in the two datasets are present for age and education. The differences in age were visible at the regional level, with the HHS dataset surveying households whose heads were upwards of ten years older than those of the households surveyed in AB-LAPOP. This difference between the two surveys is consistent over time and across countries (Figure 12).

The reason for this difference is not clear. The AB-LAPOP does not interview respondents younger than 16, while many of the National Household Surveys do include household heads who are younger than 16, although such households are rare cases. In the AB-LAPOP 2018 survey, 18 is the mode age for Colombia, Costa Rica, Dominican Republic, El Salvador, Guatemala, Honduras, and Panama.

The asterisk in the center of each graph indicates the significance level of a Wilcoxon ranked sum test between the two surveys.
The dashed vertical line represents the median of each dataset.
*Descriptive statistics for Argentina's National Household Survey are only for the urban population.
Source: AmericasBarometer 2021 and National Household Survey 2020.

Figure 12: Density plots for age of head of household show significant variation between the two datasets. Chile, Ecuador, Nicaragua, and Paraguay do not include information on age in their household survey.

Differences in education only appear at the subnational level. On average across all countries, the average level of education amounts to some secondary education, but, looking at the information by educational level we can see that the HHS dataset consistently registers a higher percentage of households as having no formal education or some primary education, whereas the AB-LAPOP consistently shows a much higher population with some secondary education. In the 2018 data the HHS data consistently calculates higher post-secondary education rates while in 2020/21 the rates are the same between surveys. A notable exception is Peru where the AB-LAPOP shows much higher post-secondary education rates.

*Discriptive statistics for Argentina's National Household Survey are only for the urban population.
Source: AmericasBarometer 2018-2019 and national Household Surveys 2018-2019.

Figure 13: Education levels accounting for survey Education levels 2020/21. There are significant differences between the representation of different education levels in the AB-LAPOP survey and the National Household Surveys. Error bars represent the 95% confidence interval.

These higher rates of representation at the lower (and higher end of the educational spectrum in 2018) in the HHS data could be due to the AB-LAPOP's much smaller sample size. Higher representation of the lower and higher ends of the educational and socioeconomic spectrum for 2018 could result in little difference in overall average of access variables, as they are generally positively and linearly correlated with income and education.

The datasets overall show similar rates of urban households; however, at the country level, this does not always hold true. Apart from Argentina, which only surveys urban families on their Household Survey, large differences in urbanization estimates between the AB-LAPOP and HHS exist for Colombia, Costa Rica, the Dominican Republic, Jamaica, El Salvador, and Uruguay.

It should be noted that the definition of rural is not uniform throughout the region.[19] In the case of the AB-LAPOP survey, the definition varies between years, with the AB-LAPOP 2019 wave aligning with the countries' census definition and 2021 employing an ordinal list with four categories: 1) a city, (2) on the outskirts or surroundings of a city/suburb, 3) in a town near a rural area/zone, and 4) in a rural area/ a rural zone. In the general analysis in this paper, 1 and 2 were considered urban while 3 and 4 were considered rural. To understand the cause of this gap, an alternative configuration was analyzed with categories 1, 2 and 3 being classified as urban while only 4 was considered rural, and the gap between the datasets narrowed but not significantly (see values in parenthesis in Table 6).

Table 6: Rates of urban households between surveys (accounting for survey design)

Urban				
Country	2018/19		2020/21	
	AB-LAPOP	HHS	AB-LAPOP	HHS
Argentina	**0.8784**	**1.0000**	**0.8683** (0.8833)	**1.0000**
Bolivia	0.6889	0.6821	0.7198 (0.7373)	0.6885
Brazil	0.8732	0.8693		
Chile			0.8754 (0.8667)	0.8881
Colombia	0.7947	0.7841	**0.6990** (0.7367)	**0.7825**
Costa Rica	**0.6338**	**0.7249**	**0.5900** (0.6169)	**0.7243**
Dominican Republic	**0.7453**	**0.8050**	**0.5840** (0.6108)	**0.8202**
Ecuador			0.7260 (0.7539)	0.7054
Guatemala	0.5060	0.4919		
Honduras	0.5358	0.5663		
Jamaica	**0.5960**	**0.5356**		
Mexico	0.7986	0.7690	0.7830 (0.8206)	0.7844
Panama	0.6985	0.7128		
Peru			0.7586 (0.8043)	0.7863
Paraguay			0.6628 (0.7132)	0.6293
Salvador	0.6255	0.6329	**0.5380** (0.5774)	**0.6241**
Uruguay	**0.9392**	**0.8441**		

Values in bold represent a difference between datasets of more than 0.05.

The differences between the datasets, specifically in their sample age and education levels likely will not have a huge impact on the water access variables at the national level, but it is valuable to be aware of these differences when working with these datasets. The variation in urban representation in some countries, however, could have significant impacts on national indicator estimates. A full breakdown of country level differences with respect to the sociodemographic variables discussed in this section is available in Annex A.

19 It is important to note that the household surveys use countries' definitions of urban and rural as implemented in the national household surveys. The JMP uses microdata developed using country definitions, but then adjusts the ratio of the population using the distribution produced by the UNPD's World Urbanization Prospects to improve international comparability.

3.2 Indicator Comparison

The three datasets discussed in this publication have different general purposes, but all include indicators that measure three common water and sanitation related concepts: **access to the piped water network, sanitary installations connected to the sewer**, and **sanitary installations connected to septic systems**. This section explores the values of these three indicators across datasets to identify situations where values differ significantly and analyze why these differences occur. Understanding the causes of these differences - whether because of sampling, methodology or definitions - allows users to determine the most appropriate dataset depending on their analytical needs.

These three indicators were chosen for comparison, because they exist in some form in all three, but also because definitions are straightforward with relatively little room for differences in methodology. Overall, the datasets are very similar, especially when the best fit lines are applied.[20] The AB-LAPOP only began incorporating water data in the 2018/19 survey wave, so the lack of data points precludes a best fit line, however generally the AB-LAPOP values fall close to the HHS survey estimates and JMP estimates. The differences between the three datasets can be attributed principally to different definitions, differences in methodology, and the previously discussed differences in microdata.[21]

There are some important definition differences to address that are common throughout the datasets before addressing the indicators one by one.

While the HHS-OLAS indicators and the indicators derived from the AB-LAPOP microdata follow many definitions standardized by the JMP, such as the concepts of improved water sources, improved sanitation facilities, and water “available when needed”, etc., the HHS-OLAS and AB-LAPOP indicators are defined in terms of residential amenities rather than service access. For example, the HHS-OLAS dataset and AB-LAPOP have distinct indicators for households with network connections vs households where the members must leave to get water from a public standpipe. In the JMP dataset, both these cases are encompassed in piped drinking water because in both the households are being serviced by piped water. This JMP indicator also encompasses other water sources such as wells and delivered water so long the household receives it via a piped system. This difference between the datasets applies for sanitation facility access as well, where the emphasis is on the type of toilet and drainage had by the home, rather than what type of toilet and drainage the home uses.

There are two reasons for this difference. The primary reason is that the phrasing on the household surveys and the AB-LAPOP survey focuses on home amenities. Most household surveys address water and sanitation in their “Vivienda” or “Home/Hogar” section, which poses questions about household amenities. Respondents are asked questions such as how many bedrooms the household has, if they have a network connection, their primary water source, the type of toilet they have, if they have a TV or computer, etc. Table 7 shows some sample questions from household surveys throughout the region and AB-LAPOP with respect to their sanitation facilities. In the context of these questions, it is much easier to harmonize indicators across surveys if they are based on household amenities, with separate categories for those who access their amenities outside the home. The JMP dataset takes a services approach which is enabled by their use of data sources outside of the national household surveys such as UNICEF’s Multiple Indicator Cluster Surveys (MICS), and is more directly relevant to SDG measurement (Table 2, Table 7).

20 The JMP Methodology uses linear regression, from which the historical indicator values are updated each year new data is incorporated (JMP, 2018).

21 This section only includes countries that have data for the three datasets included in the comparison.

*Table 7: Typical question wording on national household surveys, AB-LAPOP, MICS, and DHS (author's elaboration from household survey questionnaires throughout the region).**

Household survey	Phrasing of sanitation facility questions
2017 Encuesta Permanente de Hogares Continua (EPHC) Paraguay	ES: ¿Tiene baño? EN: Do you have a bathroom?
2018 Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) Mexico	ES: ¿Tienen excusado retrete sanitario, letrina u hoyo negro? EN: Do you have a WC, latrine or cesspit?
2018 Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) Honduras	ES: ¿Qué tipo de servicio sanitario o letrina tiene? EN: What type of sanitary service or latrine do you have?
2018 Encuesta Nacional de Empleo e Ingresos (ENEI) Guatemala	ES: ¿Qué tipo de servicio sanitario tiene este hogar? EN: What type of sanitary service does this household have?
2018 Encuesta de Hogares de Propósitos Múltiples (EHPM) El Salvador	ES: ¿Tiene servicio sanitario esta vivienda? EN: What sanitary service does this household have?
2020 Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) Ecuador	ES: ¿Con qué tipo de servicio higiénico cuenta el hogar: EN: What type of higienic service does this household have?
2018 Encuesta Nacional Continua de Fuerza de Trabajo (ENCFT) Dominican Republic	ES: ¿La vivienda posee…? [Lista de tipos de inodoros] EN: The house possesses…? [List of toilet types]
2018 Encuesta Nacional de Hogares (ENAHO) Costa Rica	ES: ¿Esta vivienda tiene servicio sanitario… [Lista de tipos de baños y conexiones] EN: This house has sanitary service via… [List of types of toilets and connections]
2017 Encuesta de Caracterización Socioeconómica Nacional (CASEN) Chile	ES: La vivienda donde usted vive, ¿dispone de sistema de eliminación de excretas? EN: Does the home where you live have a system for eliminating excreta available?
2018 Latin American Public Opinion Project, Americas Barometer	ES: ¿El baño de esta vivienda está conectado a… EN: The bathroom or toilet facility/sanitary in this household is connected to…
Multiple Indicator Cluster Surveys	ES: ¿Qué clase de instalación sanitaria utilizan por lo general los miembros de su hogar? EN: What kind of toilet facility do members of your household usually use?
Demographic and Health Surveys	ES: NA EN: What kind of toilet facility do members of your household usually use?

**Note: National household surveys and the AB-LAPOP survey ask about installations, i.e. if the household has toilet facilities and what kind. International surveys such as the MICS and DHS use the JMP recommendations and focus on service access, asking what kind of facility the members of the household generally use.*

This conceptual difference in indicator definitions is valuable. It is ideal that available international water and sanitation datasets provide information that is consistent but not duplicate. The JMP is the sector's primary international dataset, as it is used to measure the SDGs and generates population service estimates for all countries in LAC. The direct duplication of the JMP dataset would add no value. This different focus allows for a slightly different lens that addresses different policy questions than those addressed by the JMP dataset. This includes questions about public water and sewerage network coverage, the importance of public restrooms for service access, the prevalence of and reasons for bottled water use, the perception of network water quality, and the prevalence of decentralized groundwater usage and therefore the importance of groundwater quality monitoring.

This distinction between definitions is important to acknowledge when comparing indicators, although it generally has a very small impact on the end indicator value as households generally use the amenities they have in-house. The distinction, however, can be seen for households that do not have amenities. For example, the HHS-OLAS has an indicator that represents households without sanitation facilities (in home or on plot), and within this indicator are sub-indicators that represent the alternatives including

open defecation and use of a neighbor's or public facility. The JMP dataset, alternatively, categories populations that do not have a facility in-home but use neighbor's toilets or public facilities under the type of sanitation service they are using, categorizing those that report using no toilet as practicing open defecation.

3.2.1 Access to the piped water network

Piped water data is similar across datasets, although there are some notable exceptions that can be seen in Figure 14. These stem from differences in definitions, questions, and response options between the surveys.

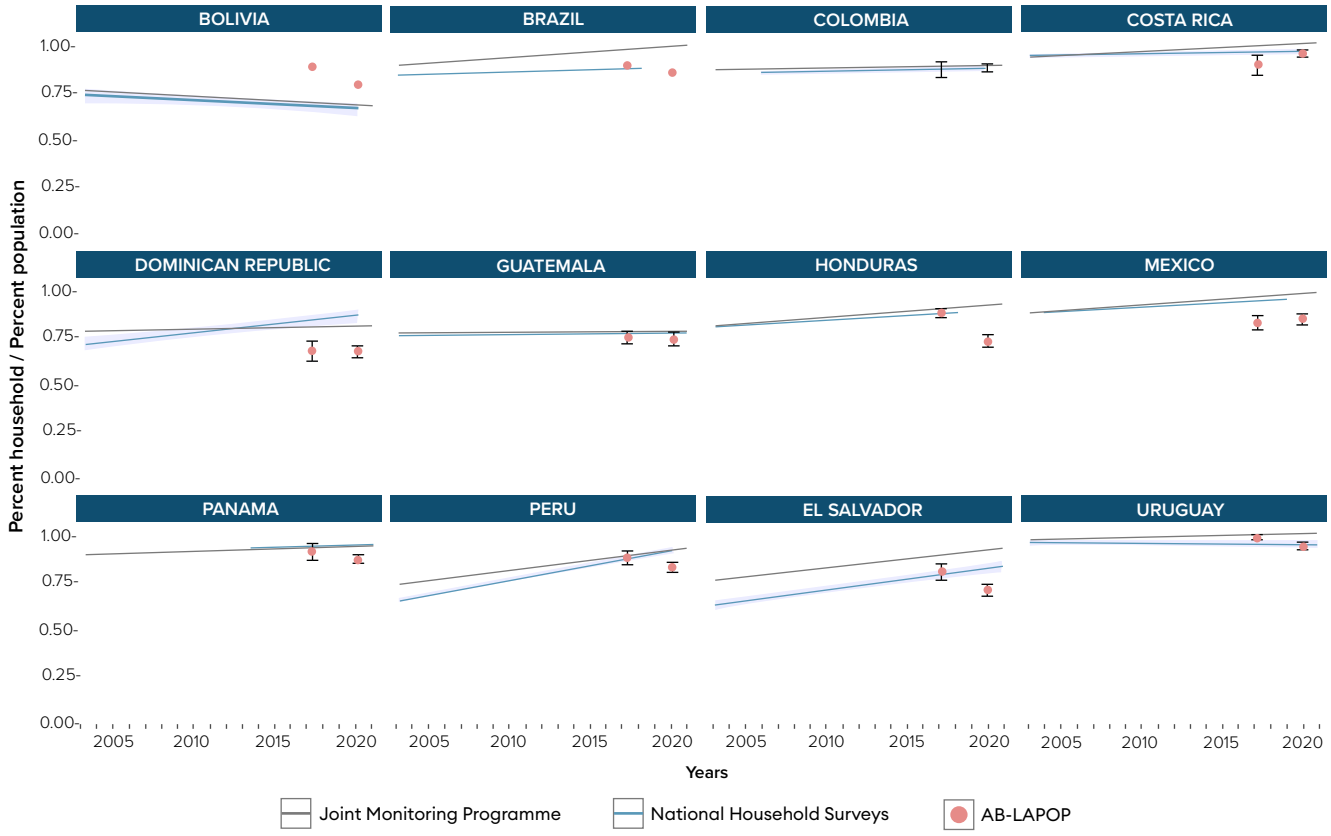


Figure 14: Rates of piped network connections across datasets

3.2.1.1 Differences in definitions

As previously discussed, the JMP defines piped water service as households with piped water, regardless of source. Households that do not have piped water connections but bring water from the public network outside the home (from a public standpipe or the neighbor's connection) are also commonly included in the calculation of this indicator. This service access definition differs from the definition in the HHS-OLAS dataset and the AB-LAPOP, which measures household network connections. This can be seen in the JMP documentation for Costa Rica, Brazil, and El Salvador. In Costa Rica and Brazil, only measuring piped water access without accounting for the source results in higher estimates in the JMP survey relative to the HHS-OLAS and AB-LAPOP estimates (Figures 15, 16), while in El Salvador the incorporation of other sources (public standpipes and neighbors' connections) result in higher estimates (Figure 17). In all cases, these differences in definition can be seen when comparing values from the datasets, where Brazil, Costa Rica, and El Salvador have higher values for piped water access when compared to values in the HHS-OLAS and AB-LAPOP datasets. (Figure 14).

While this difference may seem trivial it can have quite a significant impact on the final indicator, especially for populations that do not rely primarily on the public water network such as rural households. In Brazil, for example, 34.5% of the rural population has a connection to the public water distribution network according to the HHS dataset while, 84.7% of the rural population in Brazil has piped water (to the plot or into the home) according to the JMP dataset. This figure includes households with piped water sourced from wells, springs, rain, or other sources.

Table 8 shows the responses on Brazil's 2018 Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC) survey for questions number 7 (type of source) and 10 (source location) tabulated. While at the national level the majority of those with piped water receive it from the distribution network, a sizable proportion of respondents reported having piped water from other sources such as wells or springs.

Use of drinking water sources	**Costa Rica**			
CRI_2018_ENH	INEC			
Survey with microdata	Encuesta Nacional de Hogares, 2018			
Definitions	**Facility type estimates**	**Urban**	**Rural**	**National**
Original denomination	**Classification**	**Urban**	**Rural**	**National**
	Tap water	99.8	99.1	99.6
	House connections	99.6	98.7	99.3
Tubería dentro de la vivienda	Piped water into dwelling	98.5	95.8	97.8
Tubería fuera de la vivienda	Piped water to yard/plot	1.1	2.9	1.6
Tubería fuera de lote o edificio	Public tap, standpipe	0.3	0.4	0.3
Acueducto Rural	Other	0.0	0.0	0.0

Question		**Categories**	
no	**Description**	**no**	**Description**
1	¿Esta vivienda se abastece de agua por?	1	Tubería dentro de la vivienda
		2	Tubería fuera de la vivienda, pero dentro del lote o edificio
		3	Tubería fuera del lote o edificio
		4	No tiene por tubería
2	¿El agua que consume proviene de?	1	Un acueducto de A y A
		2	Un acueduto rural
		3	Un acueducto municipal
		4	Una empresa o cooperativa
			Un pozo
			Un río, quebrada o naciente?
			Lluvia u otro

Figure 15: JMP documentation with the definition for each category according to the data source (left), Questions from Costa Rica's ENAHO 2018. The HHS indicator represents households with access to a water distribution network at home or on their property, and therefore incorporates respondents that answered 1 or 2 to v11 and 1, 2 or 3 to v12. Use of public distribution connections is represented by a separate variable. The JMP, alternatively, classifies all piped water under this indicator regardless of the source, incorporating respondents that answer 1, 2, or 3 to v11 as well as those that answered 2 to v12. As a result, piped water under the JMP is calculated to be 98.7% of the population while the HHS value is calculated to be 95.4% of households.

Use of drinking water sources	**Brazil**			
BRA_2018_ PNAD	IBGE			
Survey with microdata	Pesquisa Nacional por Amostra de Domicilios, 2028			
Notes	Availability question relates only to piped water network, population coverage of piped water is >80% in 2018.			
Original denomination	**Classification**	**Urban**	**Rural**	**National**
	Tap water	**99.4**	**84.7**	**97.2**
	House connections	**99.4**	**84.7**	**97.2**
Canalizada em pelo menos un cômodo	Piped water into dwelling	98.2	77.6	95.3
Canalizada só na propiedade ou terreno	Piped water to yard/plot	1.1	7.1	2.0
	Public tap, standpipe			
	Other			

Quesito		**Categorias**	
no	**Descrição**	**Tipo**	**Descrição**
7	Qual é a principal forma de abastecimento de água utilizada neste domicilio?	1	Rede geral de distribuição
		2	Poço profundo ou artesiano
		3	Poço raso, freático ou cacimba
		4	Fonte ou nascente
		5	Água da chuva armazenada
		6	Outra
7A	Este domicílio tem acesso a rede geral de distribuição de água?	1	Sim
		2	Não
			Não aplicável
8	Nos últimos 30 dias, com que frequência e água proveniente de rede geral esteve disponível para este domicilio?	1	Diariamente
		2	De 4 a 6 dias na semana
		3	De 1 a 3 dias na semana
		4	Outra frequência
			Não aplicável
9	Este domicilio dispõe ou faz uso de reservatório, caixa d'água, cisterna, para	1	Sim
		2	Não
			Não aplicável
10	A água utilizada neste domicílio chega:	1	Canalizada em pelo menos un cômodo
		2	Canalizada só na propiedade ou terreno

Figure 16: JMP documentation with the definition of each category according to the data source (left), Questions from Brazil's PNADC (right).

The HHS indicator represents households with access to a water distribution network at home or on their property and so only includes respondents that answer both 1 and 2 to question 7 and 1 to question 10 when calculating this indicator. The JMP, alternatively, classifies all piped water under this indicator regardless of source, incorporating respondents that respond with answer 1, 2 or 3 on question 10. Respondents who answer that they do not have piped water (3 to question 10) then are categorized under a different indicator.

Table 8: 2018 PNADC Respondents in Brazil, Water source vs piped water location.

Respondents 2018 PNADC	Canalizada em pelo menos um cômodo	Canalizada só na propriedade ou terreno	Não canalizada
Rede geral de distribuição	328,843	7,926	0
Poço profundo ou artesiano	44,952	4,043	4,093
Poço raso, freático ou cacimba	20,085	1,652	6,251
Fonte ou nascente	15,201	1,143	2,375
Água da chuva armazenada	1,118	165	4,230
Outra (especifique)	3,689	359	6,529

Uso de fuentes de agua para consumo		El Salvador		
SLC_2019_EHPM	DIGESTYC			
Survey with microdata	Encuesta de Hogares de Propósitos Multiples, 2019			
Definitions	**Estimaciones del tipo de instalación**	**Urban**	**Rural**	**National**
Original denomination	**Classification**	**Urban**	**Rural**	**National**
	Tap water	95.3	76.4	88.1
Cañería dentro y fuera de la vivienda	Conexiones domiciliarias	88.8	68.8	80.8
	Agua entubada a la vivienda			
	Agua corriente al patio/ parcela			
Pila, chorro público o cantera	Fuentes públicas	2.4	2.6	2.5
Cañería del vecino	Otro	4.1	5.9	4.8

Figure 17: JMP documentation for El Salvador data 2019

Piped water access as defined by the JMP does not measure or estimate public network connections, it measures access to tap water in any form regardless of the location of access or source of water which aligns with the JMPs goal of measuring access as it relates to clean water for all – part of SDG 6. Policy makers and researchers who are interested in household public network connections should refer to other datasets.

3.2.1.2 Additional Response Options

The AB-LAPOP generally produces values close to those of the other two datasets, with a few exceptions. For example, AB-LAPOP piped water network access estimates for the Dominican Republic and Mexico are much lower than its JMP and HHS counterparts.

An investigation of the response options available in HHS-OLAS vs the AB-LAPOP reveals that the AB-LAPOP includes far more response options with respect to water sources than most of the national household surveys. In both Mexico and the Dominican Republic large numbers of respondents listed bottled water as the primary water source used for household activities in the AB-LAPOP survey, therefore lowering the rate of responses for piped network. Lack of this response option on the HHS-OLAS may result in respondents defaulting to the available option to which they have access, which in many cases is piped water (Table 9).[22] This underscores the importance of having consistent response options between countries, as

22 The same phenomenon can be seen in Figure 18 for Uruguay, where high rates of bottled water use for human consumption results in very low rates of network water consumption in the AB-LAPOP data, but, due to the option not being available on Uruguay's household survey for the question related to water sources used for human consumption, the HHS

the response options available can influence how respondents answer and therefore impact the comparability of the resulting data. In datasets like the HHS-OLAS and JMP, which harmonize information from various sources with different response options, this can produce comparability problems. To mitigate issues with comparability, it is important to view water source indicators and sanitation facility indicators holistically. This means that, when trying to compare between countries, users should view all water source indicators that are available to understand which sources are not present (meaning they are not addressed in the questionnaire) as well as the proportion of households that use "unclassifiable" source.

Table 9: General water sources used by respondents in the LAPOP 2018 survey.

Country	Red públi-ca/del grifo dentro de la casa	Red pública en el patio/ lote	Cone-xión irregular (cone-xión pirata) a red pública	Grifo co-munitario de uso público	Pozo entubado/ pozo perfora-ción (con bomba)	Pozo excavado cubierto/ aljibe (sin bomba)	Pozo excavado descu-bierto (sin bomba)	Pozo excavado descubierto (sin bomba)	Ma-nantial cubierto	Manantial descu-bierto	Recolec-ción de agua de lluvia	Agua em-botellada (botellón de agua)	Carreta con tanque pequeño/ tambor	Camión de agua	Río, arro-yo, canal, canales de riego
Argentina	1317	31	2	3	109	15	1	0	0	11	12	1	4	0	12
Bolivia	1111	278	3	17	43	20	14	6	33	11	12	0	32	25	22
Brazil	1214	25	1	3	108	30	2	5	1	3	3	2	2	23	21
Chile	1572	10	2	2	15	2	1	4	2	0	3	1	3	2	4
Colombia	1347	26	1	13	20	39	11	10	11	23	12	2	1	31	78
Costa Rica	1266	4	1	3	20	7	3	13	3	4	3	0	0	9	94
Dominican Republic	728	267	2	19	115	4	5	1	2	13	199	1	69	44	11
Ecuador	1359	18	3	11	54	14	2	1	3	6	16	3	13	9	5
El Salvador	964	144	11	24	53	37	24	10	18	4	18	3	9	33	54
Guatemala	1085	50	60	16	110	46	28	21	7	5	74	2	4	46	11
Honduras	1066	256	5	10	27	15	12	10	12	3	75	2	11	19	18
Jamaica	786	213	3	86	2	2	2	21	13	163	4	11	34	33	61
Mexico	1086	185	9	5	41	25	8	8	4	5	107	0	24	8	36
Nicaragua	982	112	2	35	79	100	56	38	19	0	4	1	10	18	24
Panama	1298	61	6	18	51	18	5	10	2	2	8	2	6	20	16
Paraguay	952	112	1	9	210	44	38	0	4	14	11	1	0	9	46
Peru	1150	50	3	29	40	6	1	19	4	1	4	5	19	3	51
Uruguay	1535	5	0	0	24	6	0	2	0	1	3	0	1	0	2

3.2.1.3 Differing Questions

The figures show that the AB-LAPOP water estimates for Bolivia are much higher than those for the HHS-OLAS and JMP. This difference is largely due to the data that is being compared. Bolivia's Encuesta de Hogares, which provides the data for the HHS-OLAS and JMP datasets, only asks respondents about their source for drinking water, while AB-LAPOP asks about both drinking water sources and water sources for general household use. Most of the surveys used in the JMP calculation of piped water ask about water source use in general, so the metrics graphed in Figure 14 are those meant to measure piped water access for general household use. Graphing the AB-LAPOP indicator that measures the network usage for drinking water instead of general use, we can see that the values for Bolivia from AB-LAPOP become more congruent with the other two datasets (Figure 18). Despite adjusting the indicator graphed to fit the Bolivian survey, AB-LAPOP values for Bolivia are still slightly high in comparison to the other two datasets. This could be related to the sample or survey design, but more research would be required to say definitively.

data records high rates as respondents defaulted to their second most-used source.

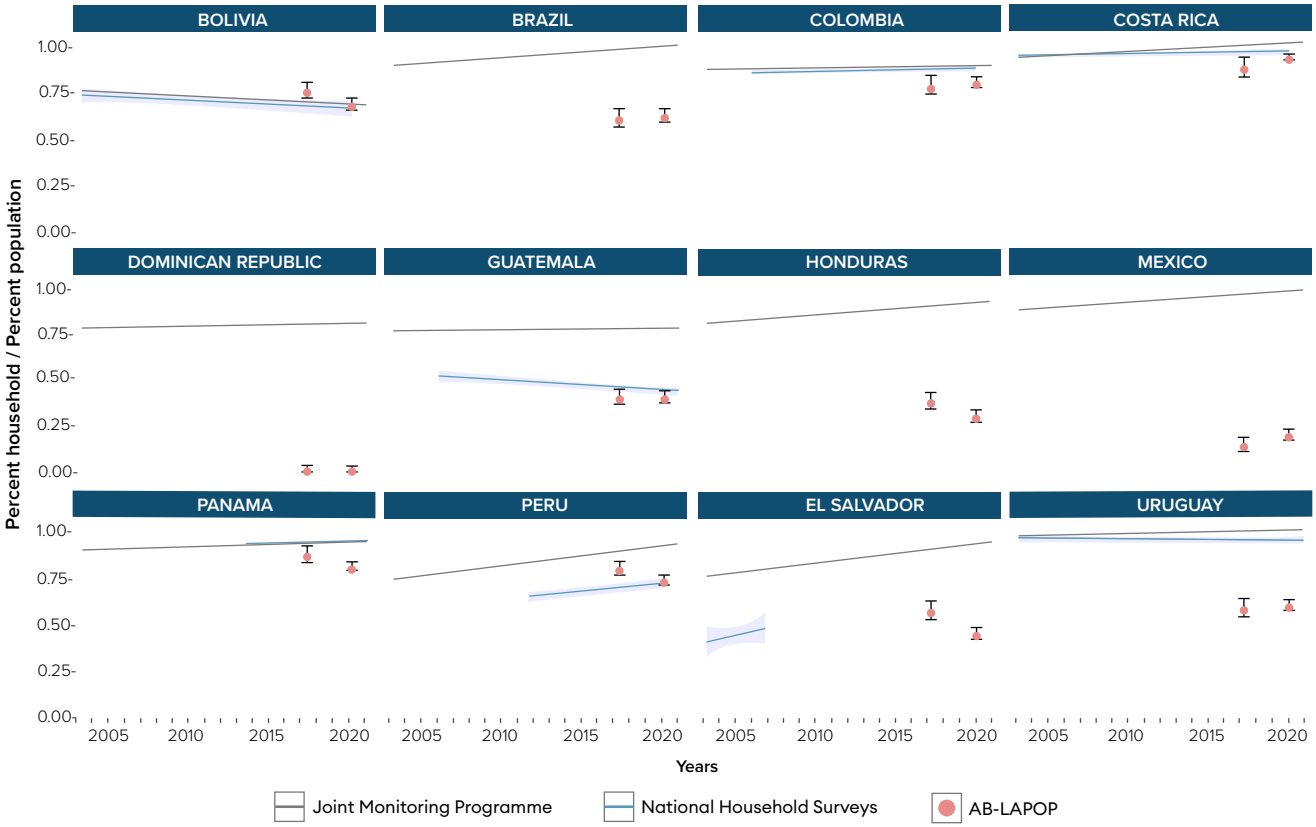


Figure 18: Piped network access in the JMP dataset, piped network household connections used for human consumption in the HHS-OLAS, and AB-LAPOP dataset. Many countries do not differentiate between drinking water sources and water, which as show above, results in large differences in values.

3.2.1.4 Sampling differences

There are some significant differences in sampling at the national level that could contribute to variations between the AB-LAPOP and the other two datasets. El Salvador, for example, has a very distinct difference in the urban sample for 2020/2021, where AB-LAPOP data shows an urban population of only 57.7% in comparison to the household survey for 2020 and 2018 as well as the AB-LAPOP 2018 which hovers around 62.5%.

Table 10: Rate of households living in urban areas between the AB-LAPOP dataset and the National Household Surveys

Country	2018		2021	
	AB-LAPOP	HHS	AB-LAPOP	HHS
Colombia	0.795	0.784	0.737	0.783
Costa Rica	0.634	0.725	0.737	0.724
Dominican Republic	0.745	0.805	0.611	0.820
El Salvador	0.626	0.633	0.577	0.624

This underestimating of the urban households could contribute to the AB-LAPOP piped water values for 2021 being lower than the HHS-OLAS values. The same pattern presents in data from Colombia and the Dominican Republic, which have lower representation of the urban households in the AB-LAPOP survey when compared to the household survey data (especially in 2021) and exhibits low AB-LAPOP estimates for piped water.

Differences in urban vs rural sampling, however, are not reflected in the piped network water indicator for all countries. Costa Rica, for example, has under representation of urban households for both 2018

and 2020, but these differences are not reflected in variation between AB-LAPOP and HHS-OLAS rates of piped network use. This could, in part, be due to high rates of piped network access among rural populations relative to El Salvador, Colombia, and the Dominican Republic (87.6% compared with 63.4%, 48.0% and 70.1% respectively (OLAS, 2023). This discrepancy in El Salvador, Colombia, and the Dominican Republic also does not have a straightforward impact on the other water and sanitation indicators examined in this section.

3.2.2 Access to the sewer network

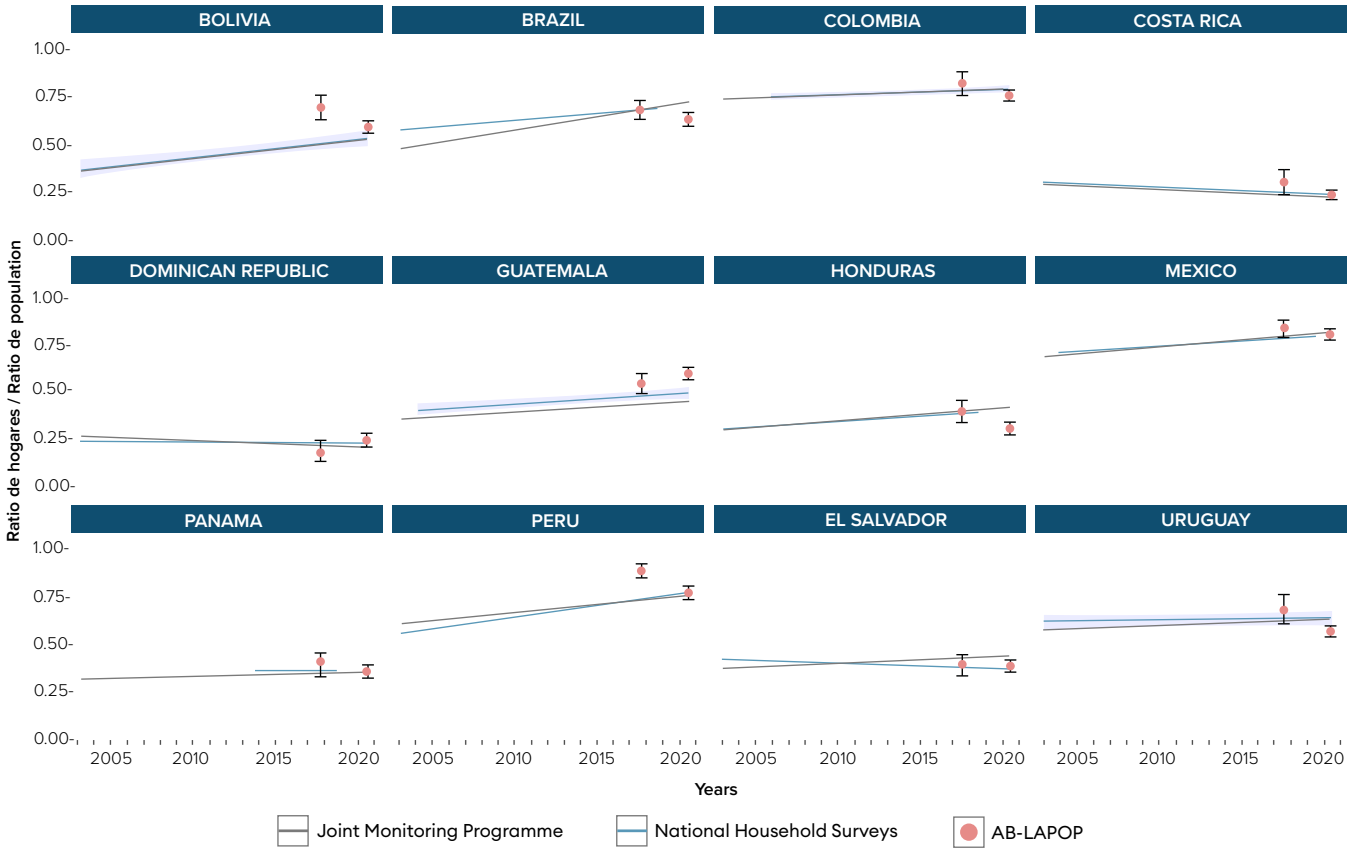


Figure 19: Sewer access rates between the JMP, HHS-OLAS and the AB-LAPOP datasets.

Sewer access is much more straightforward to measure, resulting in less differences between datasets in terms of definitions and methodology. Generally, the HHS-OLAS, JMP and AB-LAPOP datasets have very similar values for sewer network connections. One small difference between the sewer network access indicator definition between the three datasets is that the JMP dataset (and AB-LAPOP dataset in 2018) includes public and shared flush toilets in this estimate, while the HHS-OLAS includes these in other indicators. These represent a small percentage of households resulting in little change to the final indicator values.

Despite being largely the same, 2018 sewer connection estimates are high for both Bolivia and Peru. The cause of these differences is not clear. It does not appear to be linked to differences in sampling between the surveys in Brazil and Peru. Additionally, both the 2018/19 and the 2020/21 AB-LAPOP surveys directly ask about sewer access and in both Bolivia and Peru the surveys offer clear response options, although the way in which the response option is phrased varies from country to country (Table 11).

Table 11: The AB-LAPOP survey largely asks the same questions with the same response options across countries, however small differences in response options exist from country to country.

LAPOP Survey	Response option signaling the sewer network 2018	Response option signaling the sewer network 2021
Argentina	Sistema de cloacas	Inodoro conectado a red cloacal
Bolivia	Sistema de alcantarillado	Excusado/Inodoro conectado a alcantarillado
Peru	Sistema de alcantarillado o desagüe	Wáter o inodoro conectado a desagüe o alcantarilla
Guatemala	Sistema de desagüe	Inodoro conectado a drenaje
Mexico	Sistema de alcantarillado o drenaje	Excusado/taza de baño conectada a alcantarillado o drenaje

3.2.3 Septic system access

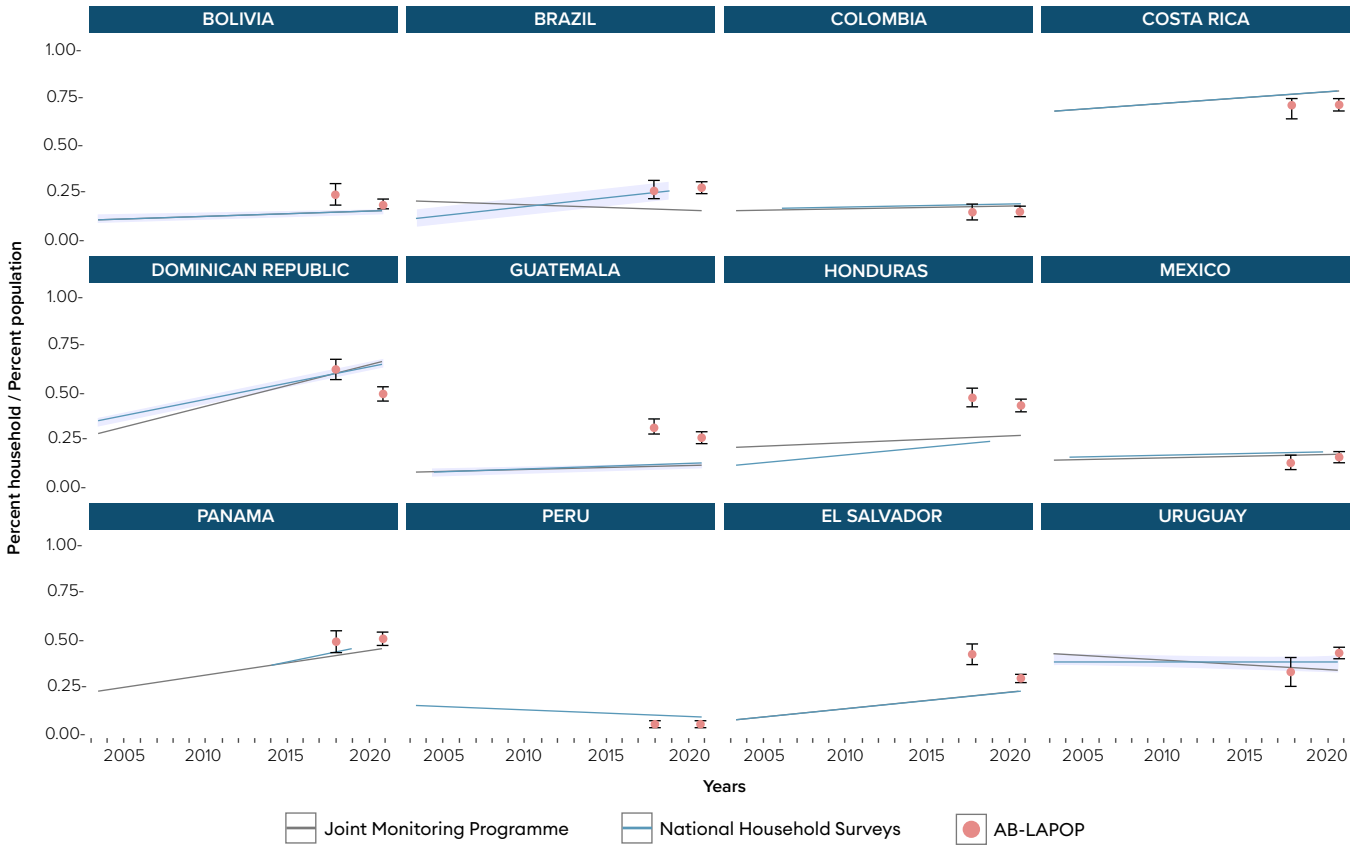


Figure 20: Septic system connection rates in the JMP, HHS, and the LAPOP datasets.

Septic system access indicators across the three datasets are also very similar, especially between the JMP and HHS-OLAS datasets. The AB-LAPOP large coincides with the other two, but has high values for Guatemala, Honduras, and El Salvador. Values are also slightly high for Bolivia. The high estimates for from the AB-LAPOP in Guatemala can be attributed to in part the wording of the response options which groups septic systems and cesspits in the same option (AB-LAPOP, 2023b). Honduras, El Salvador, and Bolivia, however, maintain separate response options, so the reason for these higher values is unknown. It could be due to differences in sampling and survey structure in the AB-LAPOP, as the smaller AB-LAPOP sample sizes result in larger confidence intervals.

Across all three indicators the values for the three datasets are similar, with some variation between countries due to slight differences in definitions and methodology. A few significant differences between the AB-LAPOP and the other two datasets for specific indicators and countries cannot be explained via differences in response options, definitions, or questions.

Sampling differences, especially differences in representation of urban populations, likely contributes to these differences. Interestingly, the values for 2018 AB-LAPOP in Bolivia are high across all three indicators, which could point to a specific issue with that survey design. These differences should be kept in mind when using the AB-LAPOP data.

It is important to note that, because there are only two years of data for the AB-LAPOP dataset, it was not possible to use a best-fit line to compare the datasets. This is important because the JMP dataset is generated via linear regression (not purely through data points generated via microdata). As a result, the information from the AB-LAPOP will likely look more congruent with the other two datasets as more years of data are obtained.

4. Conclusions

This document presents an analysis of three prominent water and sanitation datasets for Latin America and the Caribbean, detailing their differences in terms of goals, methodologies, and definitions so that users can better understand the advantages and disadvantages of using each.

The datasets compared in this document have different goals, resulting in some variation between indicators that represent similar concepts. This does not mean that one dataset is more robust than another or offers better data, but one dataset may be more *appropriate* for use depending on the users' goals. See Table 1 for more details.

Despite their different goals, methodologies, and definitions, the key indicators present similar values across datasets. There are small variations, due largely to differing methodologies related to the handling of specific cases which represent a relatively small percentage of households, resulting in small variations in the indicator values. Indicator values are especially similar between the HHS-OLAS dataset and the JMP dataset. The AB-LAPOP differs more substantially from the other two datasets but not drastically for most indicators.

There seems to be a significant difference in sampling between the AB-LAPOP dataset and the source surveys for the HHS-OLAS and JMP datasets. The AB-LAPOP tends to over-represent middle-income households and have larger margins of error due to its smaller sample size. There is also a significant discrepancy in the representation of urban vs rural populations for some countries. These differences seem to have an impact on national-level estimates, especially those with both large differences in representation and wide access gaps between rural and urban households. Despite these differences, national averages of indicators from the AB-LAPOP dataset tend to be like averages from the other data sources for indicators that linearly increase with economic prosperity, such as water network access, sewer access and septic system access. Users should exercise caution when using the AB-LAPOP for countries that exhibit discrepancies in urban/rural representation and when analyzing indicators that do not follow a linear distribution across economic strata as the averages will not conceal the selection issues. A more in-depth understanding of the impacts of these differences should be examined in future projects.

Increasing countries' capacity to collect information by improving National Statistics Systems throughout the region would benefit both national and international efforts to measure sectoral progress. The JMP and HHS-OLAS datasets both use country-generated data, which means they must contend with information harmonization and large data gaps. Improving the data collection capabilities in countries throughout the region, as well as aligning national definitions with international frameworks for data collection could greatly improve the quality of data available.

It is important to understand the methodology behind indicators to understand when best to use them. For example, a researcher trying to understand network connections should use the piped network indicators from the HHS-OLAS dataset as opposed to the JMP piped water indicator because the JMP indicator does not actually measure network connections. Alternatively, someone trying to understand access to safely managed water in the region should use the JMP as they incorporate concepts like water quality and interpolate missing data to generate representative statistics for the region. Understanding the definitions and methodologies behind these datasets is key to their proper use.

References

Canavire-Bacarreza, Gustavo; Rios-Avila, Fernando; Sacco-Capurro, Flavia, (2022). "Policy Research Working Paper 10147: Recovering Income Distribution in the Presence of Interval-Censored Data", World Bank Group, August 2022. Available: https://openknowledge.worldbank.org/server/api/core/bitstreams/f321b1be-237a-5c3d-88a3-a097c5bd9fcb/content

JMP (2018). "JMP Methodology: 2017 Update and SDG Baselines". WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. March 2018. Available: JMP methodology-Apr-2018-5.pdf (washdata.org)

JMP (2021a). "Progress on household drinking water, sanitation and hygiene 2000 2020: Five years into the SDGs." WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. September 13, 2021.

JMP (2021b). "SDG indicator metadata 6.1.1" WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. December 20, 2021. Available: https://washdata.org/sites/default/files/2022-01/jmp-2021-metadata-sdg-611.pdf

JMP (2021c). "SDG indicator metadata 6.2.1a" WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. December 20, 2021. Available: https://washdata.org/sites/default/files/2022-01/jmp-2021-metadata-sdg-621a.pdf

JMP (2023a). "Data Sources." WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. Accessed March 2023. Available: https://washdata.org/monitoring/methods/data-sources

JMP (2023b). "Downloads Index" WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. Accessed March 2023. Available: https://washdata.org/data/downloads#WLD

JMP (2023c). "Sanitation." WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. Accessed March 2023. Available: https://washdata.org/monitoring/sanitation

JMP (2023d). "SDG Monitoring." WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. Accessed March 2023. Available: https://washdata.org/how-we-work/sdg-monitoring

LAPOP (2023a). "About the AmericasBarometer." Latin American Public Opinion Project, Vanderbilt University. Nashville, USA, 2023. Accessed: March 2023. Available: https://www.vanderbilt.edu/lapop/about-americasbarometer.php

LAPOP (2023b). "Studies by Country." Latin American Public Opinion Project, Vanderbilt University. Nashville, USA, 2023. Accessed: March 2023. Available: https://www.vanderbilt.edu/lapop/

Martínez-Espiñeira, Roberto and Pérez Urdiales, María (2024). "Working Paper No IDB-WP-01571: Water Affordability Measures Under Multiple and Non-Exclusive Sources in Latin America and the Caribbean." Inter-American Development Bank, January 2024.

Montalvo, J. Daniel; Pizzolitto, Gorgina; Plutowski, Luke. "Methodological Note #009: Telephone Sampling in the 2021 AmericasBarometer". Vanderbilt University. May 24, 2022. Available: https://www.vanderbilt.edu/lapop/insights/IMN009en.pdf

OLAS (2023a). "OLAS/SCL Water and Sanitation Household Survey Dataset." Water and Sanitation Observatory for Latin America and the Caribbean. Accessed: July 2023. Available: https://mydata.iadb.org/Water-and-Sanitation/OLAS-SCL-WASH-Household-Survey-Dataset/bjat-gfsm

OLAS (2023b). "Household Survey Dataset Methodology - 2023 Update." Water and Sanitation Observatory for Latin America and the Caribbean. Accessed: July 2023. Available: https://mydata.iadb.org/Water-and-Sanitation/OLAS-SCL-WASH-Household-Survey-Dataset/bjat-gfsm

Annex A. Country-level sociodemographic representation

Topic	Country	2018/19		2020/21	
		AB-LAPOP	HHS	AB-LAPOP	HHS
Average Income (USD 2018 and 2021 respectively)*	Argentina	798	598	453	4417
	Bolivia	409	438	353	467
	Brazil	556	721		
	Chile			818	793
	Colombia	441	421	261	333
	Costa Rica	786	851	686	763
	Dominican Republic	349	447	355	329
	Ecuador			420	471
	Guatemala	312	370		
	Honduras	186	277		
	Mexico	268	472	292	413
	Panama	593	853		
	Peru			301	332
	Paraguay			381	395
	Salvador	323	317	329	301
	Uruguay	1163	1114		
Urban	Argentina	0.878	1.000	0.904	1.000
	Bolivia	0.689	0.682	0.737	0.688
	Brazil	0.873	0.869		
	Chile			0.867	0.888
	Colombia	0.795	0.784	0.737	0.783
	Costa Rica	0.634	0.725	0.617	0.724
	Dominican Republic	0.745	0.805	0.611	0.820
	Ecuador			0.754	0.705
	Guatemala	0.506	0.492		
	Honduras	0.536	0.566		
	Jamaica	0.596	0.536		
	Mexico	0.799	0.769	0.821	0.784
	Panama	0.699	0.713		
	Peru			0.804	0.786
	Paraguay			0.713	0.629
	Salvador	0.626	0.633	0.577	0.624
	Uruguay	0.939	0.844		
Age	Argentina	41.8	52.4	41.8	50.9
	Bolivia	39.7	47.5	38.7	48.8
	Brazil	39.6	49.6		
	Chile			43.1	52.7
	Colombia	40.5	48.2	40.0	48.4
	Costa Rica	40.8	51.4	40.4	53.1
	Dominican Republic	40.2	49.1	39.0	49.8
	Ecuador			37.1	51.4
	Guatemala	38.1	48.1		

Topic	Country	2018/19		2020/21	
		AB-LAPOP	HHS	AB-LAPOP	HHS
Age	Honduras	38.2	49.5		
	Jamaica	40.4	52.8		
	Mexico	42.2	49.8	42.2	51.2
	Panama	39.8	52.5		
	Peru			38.7	52.1
	Paraguay			39.5	47.4
	Salvador	40.2	50.5	39.7	50.7
	Uruguay	46.2	52.7		
Education	Argentina	3.56	3.52	2.08	2.08
	Bolivia	3.27	2.93	2.14	1.79
	Brazil	2.59	2.64		
	Chile			2.19	2.13
	Colombia	2.70	3.01	2.00	1.83
	Costa Rica	2.68	2.82	1.85	1.62
	Dominican Republic	2.99	2.81	1.79	1.75
	Ecuador			1.97	1.61
	Guatemala	2.55	2.27		
	Honduras	2.37	1.92		
	Jamaica	3.14	3.09		
	Mexico	2.87	2.89	1.89	1.80
	Panama	3.28	3.12		
	Peru			2.29	1.85
	Paraguay			1.87	1.75
	Salvador	2.68	2.21	1.79	1.42
	Uruguay	2.89	3.04		

**AB-LAPOP incomes are imputed, so an analysis of differences is not considered in this report.*