Mihály Fazekas [1], *Umrbek Allakulov* [2], *Alfredo Hernandez Sanchez* [3]

Water and Sanitation Sector Integrity Risk Index

Woking paper series: GTI-WP/2020:04
Version 2.0
February 2022, Vienna, Berlin, and Barcelona

[1] Government Transparency Institute and Central European University
[2] Water Integrity Network
[3] Government Transparency Institute and Institut Barcelona d'Estudis Internacionals

Water and Sanitation Sector Integrity Risk Index

Abstract

We employ a data-driven approach to develop a composite Water Integrity Risk Index (WIRI) made up of a host of objective proxy indicators as well as survey-based measures of corruption experience to identify and assess integrity risks in the urban W&S sector in selected urban areas around the world. Unlike broader-scope corruption indices, the WIRI outlined in this paper uses administrative datasets and survey data capturing information on corruptible transactions; thus, our analysis is micro-level, narrowly focuses on the W&S sector and is both transparent and replicable. The result is an actionable index which measures integrity risks over 9 countries between 2012 and 2021.

KEYWORDS

Corruption Risk Index, Water and Sanitation, Survey Data, Public Procurement

1. Introduction

Violations of integrity, fraud, and corruption result in reduced quality, affordability and availability of water and sanitation services. There is an urgent need to a) proactively and systematically identify, b) precisely and comprehensively measure, and c) effectively mitigate corruption risks in urban water and sanitation service provision.

The adverse effects of corruption and weak integrity on quality of life, state capacity, public services provision and economic output in the utilities sector have been widely addressed in scholarly literature (Atangana Ondoa, 2017; Chakraborty et al., 2014; Kirkpatrick et al., 2006; OECD, 2007). Acknowledging these unfavourable effects has led many international organisations and governments to call for effective action for strengthening integrity. However, policy reform effectiveness and adequate policy targeting are difficult to gauge without valid and reliable measurements of corruption.

This paper fills this gap by developing a novel measurement of integrity in the water and sanitation (W&S) sector in urban areas. It utilizes a data-driven approach to develop a composite Water Integrity Risk Index (WIRI) made up of a host of objective proxy indicators as well as survey-based measures of corruption experience to identify and assess integrity risks in the urban W&S sector in selected urban areas around the world.

The novelty of our approach comes from applying Big Data methods to administrative data and survey datasets in order to develop a comprehensive and actionable integrity risk indicator. To our knowledge, there is no integrity risk index for the W&S sector to date. Existing indexes focus on two aspects. The first is country-level reports of perception of corruption provided by sources such as the Political Risk Service, International Country Risk Guide, and

Transparency International's Global Corruption Index (Drury et al., 2009; Guasch & Straub, 2009). The second focuses on state-owned enterprises' transparency which is related to integrity but only partially overlaps with it. For example, Transparency International (TI) has developed indicators that measure the level of transparency of Public and State-Owned Enterprises based on the availability of free access to information. TI also evaluates and ranks companies based on indicators of the level of data transparency per enterprise and the legal framework to make information available (Marek Chromý, Milan Eibl, Nemanja Nenadic, Zlatko Minic, 2019). Neither of these approaches focuses on direct and measurable corruption indicators specific to the W&S sector.

By contrast, as the WIRI outlined in this paper uses administrative datasets and survey data capturing information on corruptible transactions; our analysis is micro-level and narrowly focuses on the W&S sector. In addition, this analysis rests on open data sources, making our measurements both transparent and replicable. The proposed WIRI will assist policymakers in identifying Water and Sanitation Integrity Risks which supports better policy decisions by:

- facilitating decisions about monitoring, audit, and investigations;
- informing sector-wide policy decisions for example on regulations and oversight; and
- supporting civil society and other stakeholders to hold governments accountable and advocate for better services.

The report is structured as follows; first, we outline a focused review of the literature on integrity and corruption to identify relevant actors, transactions, data sources and forms of potential wrongdoings. Next, we provide a detailed description of the methodology and we describe the criteria for selecting case studies and the resulting sample and datasets. Finally, we calculate a host of elementary risk indicators and use a set of advanced data analytic methods for parametrising and validating each of them to define the building blocks for the composite score. We present the Water Integrity Risk Index (WIRI) and review its statistical properties, contrasting urban W&S sectors across the pilot countries and urban areas.

2. Literature review

In this section, we review the relevant academic and policy literature on integrity and corruption, focusing on the W&S sector. We address the following guiding question:

What are the most important actors, transactions and forms of wrongdoing that contribute to weak integrity in the urban W&S sector?

While reviewing these concepts has its merits on its own, given the wide scope of this literature we must remain selective. As we adopt predominantly a quantitative approach, the literature review will focus on theoretical concepts and discussions which aid subsequent measurement efforts.

The literature review is elaborated through an exhaustive search of available academic literature using four sources: (1) Google Scholar (2) Scopus, the largest abstract and citation database of peer-reviewed literature (3) DiscoverEd and (4) a review of references provided by the Scopus article result, the main search terms across the sources are "Water" AND "Corruption"[1].

2.1 Understanding integrity and the lack of it

The presence of corruption or lack of integrity is a phenomenon notoriously hard to measure, partially because its definition is subject to debate (Michael, 1996). Many definitions are so broad or vague that they are not suitable for guiding measurement. For example, the OECD defines public integrity as "the consistent alignment of, and adherence to shared ethical values, principles and norms for upholding and prioritising the public interest over private interests in the public sector" (OECD, 2017). Yet this demands a definition of what public interest is and what shared ethical values are. For any measurement exercise leading to actionable and comparable results, a clear benchmark needs to be set out.

In line with recent advances in conceptualising corruption and integrity, we define integrity as the open, fair, and impartial allocation of public resources to all citizens without favouring those with connections to the detriment to outsiders without such ties (e.g. family, friendship or bribery-based) (Mungiu-Pippidi, 2006; North et al., 2009; Rothstein & Teorell, 2008). This definition is not only conceptually sound, but it also resonates with everyday understandings of integrity and lack of corruption while supporting a coherent and tractable measurement framework as it will be shown below. In addition, this definition of integrity is closely matched by encompassing definitions of corruption, hence we use lack of integrity and corruption interchangeably.

When integrity is weak, a range of corrupt activities can arise such as bribery, nepotism, theft, and other misappropriation of public resources – see (Bardhan, 1997; Nye, 1967; Lambsdorff, 1999; Shleifer & Vishny, 1993). Such corrupt acts may involve bribery and transfers of large cash amounts as kickbacks, but may also be conducted through broker firms, subcontracts, offshore companies, and bogus consultancy contracts. By implication, not everything designated as lacking integrity under this definition represents illegal activities as defined by the law in each country (Fazekas et al., 2016; Fazekas & Kocsis, 2020). Similarly, metrics of integrity in the W&S sector have been linked to efficiency and cost-effectiveness in the provision of services. Risk indicators such as the lack of more than one bidder for water-related public procurement contracts correlate with higher relative prices and a greater incidence of cancellations and delays (Fazekas et. al., 2020).

Our definition of integrity focuses attention on open and impartial access to public resources, thus allowing for a clear-cut measurement framework (Mungiu-Pippidi, 2006; North et al., 2009). However, it concerns the access to, and distribution of public resources given predefined policy goals, rather than the overall amount of such resources or the efficiency of the public sector to care for its citizens. Hence, we clearly differentiate lack of development

[1] Other keywords include: "water and sanitation" AND "Corruption"; "utility" AND "water" AND "Corruption". "Corruption" "Economy" "water".

from lack of integrity and we also separate policy making from integrity in policy implementation. These distinctions are crucial because the Water Integrity Risk Index is designed to measure the links between integrity and development without conflating the two by, for example, mixing the lack of services with the partial or biased distribution of limited available public resources among different groups.

2.2 Key actors and interactions in the Water and Sanitation Sector

This sector is best defined as the infrastructure and services related to providing safe and quality drinking water and sanitation services (Baillat, 2013; Das et al., 2016). Corrupt acts in the W&S violate the obligation to protect the human right to water (Baillat, 2013). These acts lead to arbitrary or unjustified disconnection or exclusion from water services or facilities and discriminatory or unaffordable increases in the price of water (Auriol & Blanc, 2009). Moreover, corrupt exchanges in the value chain of water utilities (inputs) also affect access to water connections and sanitation services (outputs).

The specific nature of corruption exchanges in the W&S sector is largely due to the constellation of actors, their typical interactions, and structural constraints and enablers of corruption such as a monopoly provider position. The actors interact on different levels in the sector: country level, urban area level, provider level and project level (Halpern et al., 2018). In the W&S sector, the literature identifies the following key actors[2] (Davis, 2004; Punjabi, 2017):

- the customers,
- the staff of the local utility: professional, engineering staff as well as senior administrators,
- political, bureaucratic leaders, and regulators
- contractors.

Given the different sets of public and private actors in the W&S sector (Jergelind, 2015), corruption can take various forms depending on the underlying interactions and structures. Hence, we define two levels or types of corrupt violations of integrity: grand corruption and petty corruption.

First, grand corruption in the sector is defined as bribes, kickbacks, or any other favour received by politicians, civil servants or utility leadership to give undue support or to award contracts to selected consultancy firms, constructing firms, and additional water and other sanitation-related companies. Another element that defines high-level corruption in the water sector is that companies create grand corruption networks through political groups and alliances with local and international actors which create an oligarchy in order to control the market and block competition (Hall & Lobina, 2007). Specific actors in a grand corruption scheme often also include multinational and local construction companies who win

[2] These actors are defined based on the regulatory and organisational context of local water utilities as well as detailed qualitative research (Davis, 2004).

engineering and public works projects (Hall & Lobina, 2007). Importantly from the perspective of corruption, the sector is extremely concentrated. For example, there are only four major companies operating in Sub-Saharan Africa: Thames Water, Vivendi, ONDEO, and SAUR, and two more at the global level: AngliaWater and YorkshireWater (Auriol & Blanc, 2009).

Government contracting and the capture of government policy by elites are particularly prevalent in low-integrity settings. Private sector firms and the lucrative service, construction, and public-private partnership (PPP) contracts they receive represent a major channel for siphoning off public funds in low-income settings such as Sub-Saharan African countries (Auriol & Blanc, 2009). An example of such a scheme was revealed by the prosecution for bribery of 19 international construction and consultancy firms in Lesotho Highlands Water Project (Earle, 2007).

A frequently quoted scheme of grand corruption involves dubious privatisation which lends control of end-user prices to the involved corrupt network typically consisting of private entrepreneurs and politicians (Auriol & Blanc, 2009). Keeping end-user prices high and hence earning corrupt rents are enabled by the monopoly position of the utility company (Auriol & Blanc, 2009). Nevertheless, public ownership can also enable corruption where regulations stipulate controlled prices generating large profits at the utility which then subsequently are siphoned off through subcontracts, wages and outright stealing.

Second, petty corruption in the W&S sector involves cash bribes from customers to low or middle-level civil servants to facilitate or speed up the delivery of W&S services (Rafi et al., 2012). Customers can be categorized into two groups (1) individual residential clients and (2) executive clients that have economic activity in the industry (for example, company owners, entrepreneurs, businesspeople). Bribery schemes are identified through surveys where low civil or middle civil servants admit to either demanding bribes or users admit to supplying bribes in order to receive the service or improve the services they receive. Mapping key actors, their institutional roles, and the interactions among them contributes to our understanding of integrity violations.

Based on the above, our framework focuses on 3 main pillars of integrity in the W&S sector:

1. Public investment projects (e.g., building new pipelines or drainage),
2. Recurrent spending supporting ongoing operations (e.g., paying salaries, purchasing computers), which is addressed as operations in this work; and
3. Client-utility interactions (e.g., paying utility bills).

Violations to integrity in the first area clearly falls in the domain of grand corruption, while violations in the third area typically involve petty corruption. We also differentiate the second area because it captures the internal processes of the W&S services provider unaccounted for by the two other areas. In this area, the violations to integrity can relate to both grand and petty corruption.

2.2.1 Corruption in investment projects

Corruption in investment projects in the W&S sector typically ends up happening through public procurement or government contracting. In public procurement, the aim of institutionalised corruption is to steer the contract to the favoured bidder without detection in a recurrent and organised fashion (Fazekas & Tóth, 2014; World Bank, 2009). Corruption in public procurement requires at least two violations of principles of fair distribution of public resources: 1) avoiding competition, by for example using unjustified sole-sourcing or direct contract awards; and 2) favouring a particular bidder, by for instance tailoring specifications, or sharing inside information. This definition of corruption focuses attention on restricted access to and unfair competition for public resources (Mungiu-Pippidi, 2014; North et al., 2009).

Often, contractors compete against each other by partnering with elected officials and senior bureaucrats who can provide insider information and/or carefully manipulate tender documents to subvert competition (Davis, 2004). Even when there is some form of competitive bidding, bidders often form cartels to set prices and who wins which contract (Davis, 2004). Furthermore, corrupt companies continue to increase their profit margins by colluding with the technical staff during the contract implementation phase (Davis, 2004). The technical staff are motivated to ensure continuation of these relationships in order to secure a steady flow of bribes.

Corrupt acts which influence bids or contracts result in fraud as over or under-valued assets. This impacts the quality of the work and the time it is completed. Additionally, fraud in invoicing may be present through marked-up pricing, and or overbilling by suppliers. This may result in not building to specification, concealing substandard work or the failure to complete works, or in the management of the service (OECD, 2007).

2.2.2 Corruption in recurrent spending supporting operations

Once the utility is operating, integrity may be lacking throughout the maintenance of the service and the execution of its budget (Plummer & Cross, 2007). This can manifest itself as administrative corruption in personnel management when presents and payments are made by candidates to receive appointment, promotion, or conserve a strategic post (e.g. utility directorships). Also, the inflated cost of the service facilitates nepotism in the hiring of technical staff (Pusok, 2016) who aim to conserve their posts in order to continue asking for bribes from the same group of people (Punjabi, 2017). These power relations allow a particular group of people to gain and maintain control of the service while continuing to undermine integrity.

In addition to nepotism, corruption in operations can take further forms. For example, senior agency administrators may ask for a payment from professional and engineering staff in exchange for favourable reviews, promotions, and transfers (Punjabi, 2017). Ghost employees on the payroll may be present (Levy, 2007) this is a practice used to pay back favours between actors. These practices lead to inadequate recruitment of staff which impacts operations. Additionally, inflated costs for the maintenance of the service relating to chemicals, vehicles or equipment are also present throughout the W&S sector (OECD, 2017). Just like for the previous area of corruption, we outline an example of how this might look like in practice.

2.2.3 Corruption in client-utility interactions

Lack of integrity in the client-utility nexus can take a variety of forms with different effects such as unaccounted for water, unofficial usage of tankers, low reporting of faults, unexplained zonal variations, and ignored complaints from consumers and small-scale providers (Gulati & Rao, 2007). Each of these outcomes of low-level corruption typically results from bribes paid by the client, private household, or company, to low-level bureaucrats of the utility company. Nevertheless, mid to high-ranking officials in the utility company may also support or even facilitate such a scheme in order to further extract rents for themselves or simply to keep under-paid bureaucrats at bay.

Payments are made in exchange for several services, such as expediting applications for new connections; quick attention to water supply works and sewer repair work; the falsification of water bills; and ignoring illegal service connections. This also impacts on a range of businesses processes as industrial actors require water to produce goods or to provide their services (Makoni, 2014).

3. Methodology

The Water Integrity Risk Index (WIRI) uses administrative datasets and survey data in order to develop a comprehensive and actionable composite index which is comparable across different organisations and over time. We identify three pillars of integrity in three areas where wrongdoing can happen: a) investment, b) operations, and c) client-utility interactions. Each of the three pillars can be assessed using a host of tried corruption and integrity indicators based on both administrative and survey data sources, resulting in a robust and comprehensive measurement.

Given that integrity is a latent variable, we must rely on proxy indicators which can, in conjunction, reveal integrity risks. The most widely used methods for latent variable estimation are principal-component analysis and structural equation modelling (Dillon et al., 1996; Hoyle, 2012; Pituch, 2015). These are widely tested and suitable methods for our purposes; however, given the small sample size and large number of missing values in our dataset, we opt for a simpler approach by generating the composite WIRI in the following steps:

1. We standardize each component indicator of integrity-risk so that they can be directly compared (higher values imply higher integrity).
2. We calculate the weight of each component indicator (five in total, categorized into 3 pillars) by the amount of data points available for the 2021 pilot timeseries (2012-2019). Fewer available data points in a component lead to a decrease its pillar weight on the index.
3. We calculate the weighted mean of each indicator (see Table 4) to derive the composite WIRI score based on the data available.

3.1 Data sources and sample

In order to identify suitable datasets and indicators, we carried out a comprehensive search strategy starting from as broad a list of countries as possible then subsequently narrowing down the list to countries and urban areas where multiple datasets and integrity indicators intersect. The search strategy focused on open sources which provide valid measures of integrity and offer a consistent dataset across time, covering the 2005-2019 period. We mapped available data sources and relevant indicators, in particular: their location and accessibility, exact definition, targeted geographical unit, time-period covered, and sector-of-measurement. Where it was needed, we requested micro-data on top of publicly available aggregates. The mapping concentrated 4 distinct types of data:

1. Surveys of corruption experiences,
2. Public procurement data, including risk indicators,
3. W&S utility data,
4. National Statistical Office data.

First, we reviewed all available cross-country surveys which enquire on corruption and integrity, specific to the water sector. This review included all available surveys from reputable sources (such as Transparency International, WHO, World Bank, and different social surveys conducted by universities and research organisations). We focused on surveys which (1) provided local identifiers, (2) specifically covered water sector corruption, and (3) asked about direct experiences with corruption. This filtered out surveys which focus on the perception of corruption or provide country level aggregates. The full list can be consulted in Appendix A. Reviewed Sources for Corruption In W&S.

Second, we checked the list of countries with suitable survey data against the list of countries where corruption proxy indicators were readily available for the research team. The corruption proxy indicators are represented by public procurement datasets which have been collected by the Government Transparency Institute from official government data repositories and publication portals.[3]

In the 2020 pilot of the index, countries and their urban areas were identified based on the scope, quality, and availability of data per year in all relevant datasets. In addition, our aim was to offer a global sample of countries including countries from as many continents as possible. In this iteration of the WIRI, we have added two countries – Bangladesh and Peru – as well as 3 additional urban areas in the case of Kenya. As of 2021, the WIRI has been implemented in 34 cities in the following countries:

1. Georgia
2. Hungary
3. Kenya
4. Paraguay
5. Romania
6. Uganda
7. Uruguay
8. *Peru*
9. *Bangladesh*

We construct a tailored list of keywords for each urban area in order to identify each exclusive interaction by the W&S sector in each city; the interaction is represented by inputs purchased by utilities (e.g. office supplies or pipes) or outputs provided by them (e.g. water services to government ministries). We identified the relevant contracts either by searching for the utilities'

3 See: https://public.tableau.com/profile/mihaly.fazekas#!/vizhome/GTIDataScope/GTI_DataScope?publish=yes

names in the buyer name field of the public procurement datasets; or by delimiting product codes and names specific to the W&S sector (wi

Appendix B. Keywords for searches in public procurement data) Below is a summary of available tender contracts per urban area.

Due to the different data sources for each variable, we utilised merging techniques in order to create a single data set. Among the countries where both survey and public procurement data were available, and in order to account for Water Utility Risk, we further looked into water utility companies' annual reports and websites in search for declared expenses on different materials, employee salaries and yearly revenue. The addition of utilities required mapping how the utility interacted with the government and possible auditing agencies or Water Ministries that contain financial reports. Additionally, we screened national statistical offices for relevant information to the W&S sector, this includes indicators relevant to a country's W&S infrastructure.[4]

Though we attempted to select urban areas with a comprehensive data coverage, even this shortlist included several urban areas where not all identified data sources were available for every year. For example, in most cases, the survey of corruption experiences was run 2 or 3 times in the last 10 years. We report the sparsity of data points between 2012 and 2019 in Table 2. The table summarises the component indicators (described in section 4.2), the missing data points for each indicator in a balanced panel dataset (urban areas and years), the rate of data availability and the calculated weight for each component based on the former.

TABLE 1. DATA SPARSITY PER INDICATOR (2012-2019)

Variable	Pillar	Missing Data	Available Data Rate
avg_cri_inv_int_100	Investment Risk	10	0.89
pipe_int	Investment Risk	63	0.34
avg_cri_op_int_100	Operations Risk	11	0.88
avg_cri_inter_int_100	Client-Utility Interaction Risk	58	0.39
cui_survey_int	Client-Utility Interaction Risk	92	0.04

In this new iteration (2021) of the WIRI, we keep the same pillar weights as in the 2020 pilot study. This largely reflects the lack of availability of survey data and pipe length data. We expect that soon, as more data points become available and more urban areas are included, the relative weights of each of the five source indicators will be modified.

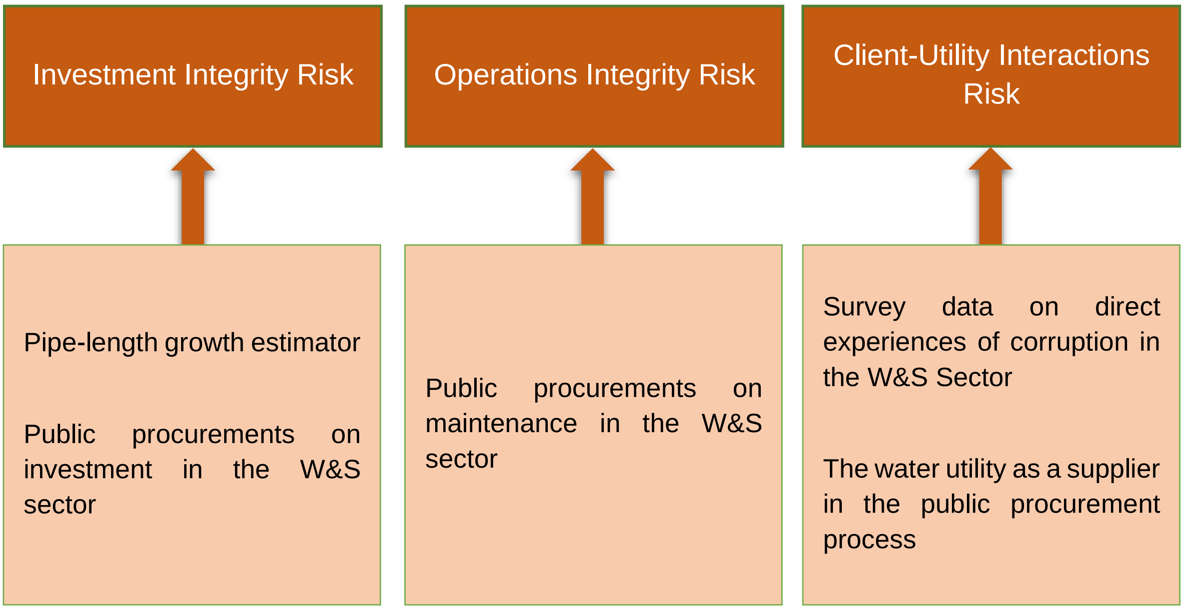
3.2 Indicators

In this section, we define each indicator and asses its strengths and weaknesses, as well as the additional indicators which would further enhance the reliability and validity of the index in

[4] In order to normalize and harmonize indicators (e.g., prices of contracts) we include a background variable: Purchasing Power Parity (PPP) conversion factor on private consumption of Local Currency Unit (LCU) per International $ provided by the World Bank.

subsequent iterations. First, we present public procurement indicators which are used in all 3 pillars and then we discuss the specific indicators for each pillar in turn.

FIGURE 1. WIRI INDEX COMPONENTS



Note: Public procurement indicators are derived from red flags in contracts following the Corruption Risk Index (CRI) (Fazekas *et. al.* 2006)

Public procurement risk indicators – which are present in the three different pillars (Investment Risk, Operations Risk, and Client-Utility Interaction Risk) – capture the risk of deliberate restrictions to open and fair competition in public tenders and contracting decisions in order to benefit a connected bidder (Fazekas & Kocsis, 2020)[5]. We assign each public procurement contract to one of the 3 pillars (investment, maintenance and client utility interaction) using product codes specific to the nature of W&S activity defined by public procurement data systems such as the Common Procurement Vocabulary (CPV) codes and the United Nations Standard Products and Services Code (UNSPSC). The data is collected using countries' national public procurement portals, thus there is always a risk of biased or invalid information being fed to those portals. However, we undergo validity checks to make sure the data is consistent to maximize reliability.

The public procurement risk indicator is a composite score of five elementary risk indicators (

Table): the length of the tendering decision period, the procedure type used to award a tender, whether there was only a single bidder for a contract, the length of the advertisement of the tender and whether the call for tenders was openly published. For ease of interpretation, we average over these 5 indicators to arrive at a composite score and use the same score

[5] Fazekas & Kocsis (2020) define the Corruption Risk Index (CRI) which denotes higher values as high corruption. We build on their work but reverse the scale to match the integrity logic of the WIRI.

calculation methodology in each of the three pillars. The composite score is scaled so that it falls between 0 and 100, with 100 representing the highest integrity and 0 representing the lowest integrity (lack of integrity). We construct weights which utilise the number of contracts in order to account for the differences between urban areas in accordance to the amount of micro interactions.

TABLE 2. PUBLIC PROCUREMENT RISK INDICATOR DEFINITIONS

Indicator name	Indicator definition
Length of decision period	***100=LENGTH OF DECISION PERIOD IS UNRELATED TO CORRUPTION RISKS (SINGLE BIDDING)*** ***0=LENGTH OF DECISION PERIOD OR MISSING DECISION PERIOD IS RELATED TO CORRUPTION RISKS (SINGLE BIDDING)***
Procedure type	***100=OPEN*** ***0=NON-OPEN (ACCELERATED, RESTRICTED, AWARD WITHOUT PUBLICATION, NEGOTIATED, TENDER WITHOUT COMPETITION)***
Single bidder contract	***100=MORE THAN 1 BID RECEIVED*** ***0=1 BID RECEIVED***
Call for tenders publication	***100=CALL FOR TENDER PUBLISHED IN OFFICIAL JOURNAL*** ***0=NO CALL FOR TENDER PUBLISHED IN OFFICIAL JOURNAL***
Length of advertisement period	***100=LENGTH OF ADVERTISEMENT PERIOD IS UNRELATED TO CORRUPTION RISKS (SINGLE BIDDING)*** ***0=LENGTH OF ADVERTISEMENT PERIOD OR MISSING ADVERTISEMENT PERIOD IS RELATED TO CORRUPTION RISKS (SINGLE BIDDING)***

The first pillar, **Investment Integrity Risk** (IIR), estimates integrity risks in investment projects, it incorporates public procurement risk indicators and a pipe length-based indicator from national statistical offices. Following existing literature (Klašnja, 2017), we posit that large investments into piping infrastructure without a corresponding increases in pipe length is of concern, thus a risk factor. We incorporate this risk indicator of missing infrastructure by comparing the total length of the network with prior investment. In a regression set-up, this indicator is defined as the error term of the panel regression regressing the change in the stock of pipe length on the current and last year's infrastructure investment value while controlling for baseline pipe network length. Pipe length is measured as the length of the total network in a urban area in kilometres provided by statistical offices in a yearly and consistent manner. In order to account for pipe investment in the model we select different pipe investment related categories from the selected W&S tenders, some examples include "irrigation, pipe construction work, bends, pipelines". We observe the missing infrastructure indicated by lower values of the residuals from the regression model which are normalized between 0 and 100.

It is important to mention that these two indicators do not map the different stages of the investment process (like the example in Pakistan, presented by Rafi, Lodi and Hasan (2012)) because of the difficulty of getting sufficiently detailed data on project stages, in particular project implementation data.

The **Operations Integrity Risk** (OIR) indicator considers the lack of integrity throughout the maintenance and operations of the service provided by the utility (Plummer & Cross, 2007). The OIR utilises the public procurement risk indicators from maintenance, as categorized by

CPV and UNSPSC codes. Examples of these include chemicals products, transportation equipment, laboratory materials, IT services.

This indicator is consistent and reliable across different years. However, it is important to clarify that the OIR indicator does not incorporate the total salary of the staff in the utility or manipulation of hiring and promotion which can result in lack of integrity (Punjabi, 2017). The following observations are made in order to integrate such data:

Observation to integrate corruption in hiring and employees in W&S

We find that in order to incorporate in OIR an indicator of personnel management it is necessary to acquire consistent data on the average salary of employees divided by different categories of types of employees ranging from technical staff to middle and high management. The differences in salaries depending on the distribution may be an indicator of lack of integrity. Unfortunately, currently available data from some of the water utilities include yearly payment of all staff, and different categories do not segregate this.

The third pillar corresponding to the **Client Utility Interaction Integrity Risk** indicator includes two metrics: a) the public procurement risk indicator and b) an indicator which integrates direct experience with corruption represented as admission of bribery by households towards the W&S service (Rafi et al., 2012, Punjabi, 2017).[6] We construct the risk component of client utility interaction integrity from the public procurement risk indicator using the water utility as a supplier in the public procurement process.

We rely on survey data from two sources to construct our metric on experience with corruption in the W&S sector (Davis, 2004; Makoni, 2014). From the Global Corruption Barometer we obtain admissions of bribery in the W&S sector for 2016. The second survey selected is the Afro-barometer. We collect positive responses from a representative sample of the population in Africa who admit to bribing to obtain water services. This survey is conducted approximately every two years. Here, we select positive answers as a response rate to the question[7]: *"And how often, if ever, did you have to pay a bribe, give a gift, or do a favour for a government official in order to get the services you needed?"*. The frequency of "once or twice", "a few times", or "often" responses is recoded as an admission of bribery.

A discussed in the literature review, individuals in surveys do not always openly disclose participating in bribery (Davis, 2004); this may result in low admission rates in urban areas. For each of the available surveys, we calculate the rate of admitted bribery by dividing the

[6] Though the index would be strengthened by combining the perspectives of corruption in the W&S sector from both households and businesses, surveys on the latter have not yielded data suitable for our purposes. An example of this is the World Bank Business Enterprise Service Modules Survey (BEEPS), which collects admissions of bribery for water and sanitation services by member of the business community. Though geolocated and topical, the number of respondents per urban area per year is very low (under 20) which is why we exclude it from this iteration.

[7] Question is branched from *"In the past 12 months have you tried to get water, sanitation or electric services from government?"* If the respondent answers *Yes*, the follow up question is asked. Survey question codes change over time. In round 7 (2019), the question is expanded to include electricity alongside water and sanitation.

number of respondents who admitted bribery over the number of respondents who required or requested a W&S service in a urban area.[8]

These surveys are used because they directly ask about paying bribes in receiving the service from water utilities as opposed to other surveys which focus on how corruption is perceived. Additionally, these are the only two surveys which contain a urban area identifier and not just country-level aggregates (such as the UN Database, GLASS, SD6 surveys, etc.). Even though both surveys come from reputable sources, the country selection and year of survey application are not always systematic.

The two surveys include admissions of bribery towards public officials, though fail to include customer bribes to falsify meter readings, the existence of illegal connections, or speed money to expedite repairs (Punjabi, 2017). To our knowledge, there is no survey that investigates these issues in a consistent, reliabile and valid manner across countries.

Given these limitations, the survey component of the WIRI index has the lowest weight (1.6%). It is important to note, however, that as survey data becomes more systematically available, the relative weight of this component could be scaled upwards in subsequent iterations (see table 2).

Finally, in our analysis of price setting as an indicator for corruption we came to the following observation:

Observation of price setting.

To measure whether corruption increases the price of water (Auriol & Blanc, 2009), we analyse the available information of cost of water for clients provided by International Benchmarking Network for W&S Utilities (IBNET). It contains more data points across time in comparison to utility reported water costs which only cover the current year. However, the measures of the cost of water have significant time gaps between them (some of up to 5 years). After careful consideration, we find that reported costs of water are insufficient to incorporate price-setting into WIRI because the variance of the cost of water can be attributed to other factors (policy changes or availability of the resource, for example), especially when there is a significant time period between the data points.

We integrate control variables to account for the differences between urban areas and the public procurement indicators, making these units relatable in context. For public procurement risk indicators, we include total number of contracts as frequency weights in the W&S sector. Equally, for the client utility interaction survey indicators we utilise the total number of respondents that required a service or a new connection as the frequency weight for the sample.

In order to normalize and provide a comparable measure, the total value of contracts is represented in International USD (GK$). The calculation uses purchasing power parity ratio provided by the World Bank which is a standard measure of price level differences across

[8] Only explicit answers are considered, non-respondents (NA) or "don't know" answers are dropped. This bribery survey metric is expressed as a percentage [0:100] where 100 means that all applicable respondents admitted to bribery for W&S services in over the relevant time-period.

countries in consumption in local currency after inflation. This is used as opposed to the GDP because it provides a universal currency based on actual prices.

The final dataset which we used for the analysis also includes a range of calculated and auxiliary variables. The full variable list, definitions, and sources can be found in Appendix C. Variable Dictionary

4. Analysis

We present the WIRI index data per urban area as both cross-section and time-series. Given the availability of surveys and investment indicators we selected a 9-year period (2012 to 2021) for both the cross-sectional and time-series analysis.

The cross-sectional composite WIRI is created based on information on all 3 pillars. For the investment risk pillar, we average over integrity risks in public procurement tenders as well as missing pipe stock. For the operations risk pillar, we could only make use of public procurement-based indicators. And for the client-utility interactions we combined public procurement-based risk scores with survey-based metrics.

Based on the three pillars discussed in the previous sections, we calculate the weighted average per urban area, where the weight of the pillars is directly proportional to the ratio of available data for each. The global pillar weights and indicator averages per urban area are summarised in Table 3.

TABLE 3. COMPOSITE WATER INTEGRITY RISK PER URBAN AREA (2012-2021)

Indicator:		*Investment Risk*		*Operations Risk*	*Interactions Risk*		**Composite Index**			
Weights		**0.35**	**0.13**	**0.35**	**0.15**	**0.02**	**48%**	**35%**	**17%**	**100 %**
Urban Area	**Country**	*Procurement Investment*	*Pipe Length*	*Procurement Operations*	*Procurement Interactions*	*Survey Data*	**WIRI INV**	**WIRI OP**	**WIRI CUI**	**WIRI**
Asuncion	Paraguay	71	NA	68	NA	62	71	68	62	69
Batumi	Georgia	61	NA	58	59	NA	61	58	59	60
Bucharest	Romania	66	27	76	44	NA	47	76	44	56
Budapest	Hungary	72	24	73	77	NA	48	73	77	62
Cluj	Romania	69	30	76	NA	NA	50	76	0	50
Gyor	Hungary	71	26	71	70	NA	48	71	70	60
Iasi	Romania	68	29	79	NA	NA	49	79	0	51
Kampala	Uganda	53	NA	53	51	NA	53	53	51	53
Montevideo	Uruguay	41	NA	42	NA	89	41	42	89	50
Nyiregyhaza	Hungary	69	28	68	69	NA	48	68	69	59
Tblisi	Georgia	66	NA	63	50	NA	66	63	50	62
Chandpur	Bangladesh	64	NA	31	NA	NA	64	31	0	42

Chattogram	Bangladesh	67	NA	50		71	NA	67	50	71	62
Cumilla	Bangladesh	62	NA	33		58	NA	62	33	58	51
Dhaka	Bangladesh	61	NA	55		72	84	61	55	36	55
Khulna	Bangladesh	64	NA	50		78	NA	64	50	78	61
Noakhali	Bangladesh	58	NA	15		63	NA	58	15	63	44
Pabna	Bangladesh	67	NA	63		75	NA	67	63	75	67
Rajshahi	Bangladesh	63	NA	60		57	NA	63	60	57	61
Rangpur	Bangladesh	72	NA	46		53	NA	72	46	53	60
Sylhet	Bangladesh	67	NA	59		67	NA	67	59	67	64
Mombasa	Kenya	65	NA	63		70	NA	65	63	70	65
Nairobi	Kenya	64	NA	59		66	80	64	59	33	57
Nyeri	Kenya	60	NA	70	NA		NA	60	70	0	53
Siaya	Kenya	64	NA	60	NA		NA	64	60	0	52
Arequipa	Peru	85	NA	87		81	NA	85	87	81	85
Ayacucho	Peru	85	NA	85		84	NA	85	85	84	85
Bagua	Peru	84	NA	84		88	NA	84	84	88	85
Cusco	Peru	88	NA	88		86	NA	88	88	86	88
El Agustino	Peru	83	NA	85		81	NA	83	85	81	83
Jesus Maria	Peru	84	NA	85		84	NA	84	85	84	84
Lima	Peru	82	NA	83		81	NA	82	83	81	82
San Isidro	Peru	87	NA	85		81	NA	87	85	81	85
Trujillo	Peru	82	NA	85		79	NA	82	85	79	83
Yurimaguas	Peru	81	NA	83		79	NA	81	83	79	82

FIGURE 2. WIRI RANKING OF URBAN AREAS (CROSS-SECTIONAL)

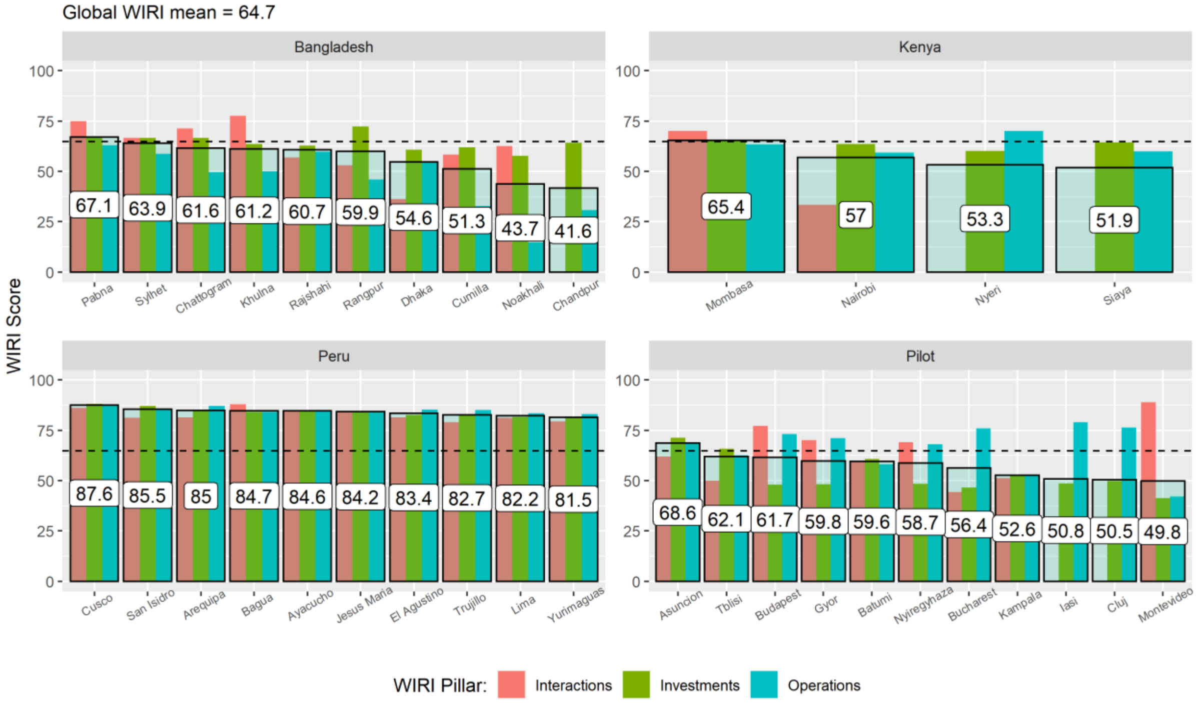


Figure 2 shows the results for the original pilot urban areas, as well as for new cities in Peru, Kenya, and Bangladesh. The pilot cities were recalculated based on two methodological changes: a) when information on one pillar is missing, the pillar is penalized as a 0, and b) when two pillars are missing, the observation in dropped completely. This was applied both to the cross-section and to the timeseries. Notably, the scores for the two cities from Romania, Iasi and Cluj Napoca, have decreased since there is no data available for those cities in the Interactions pillar.

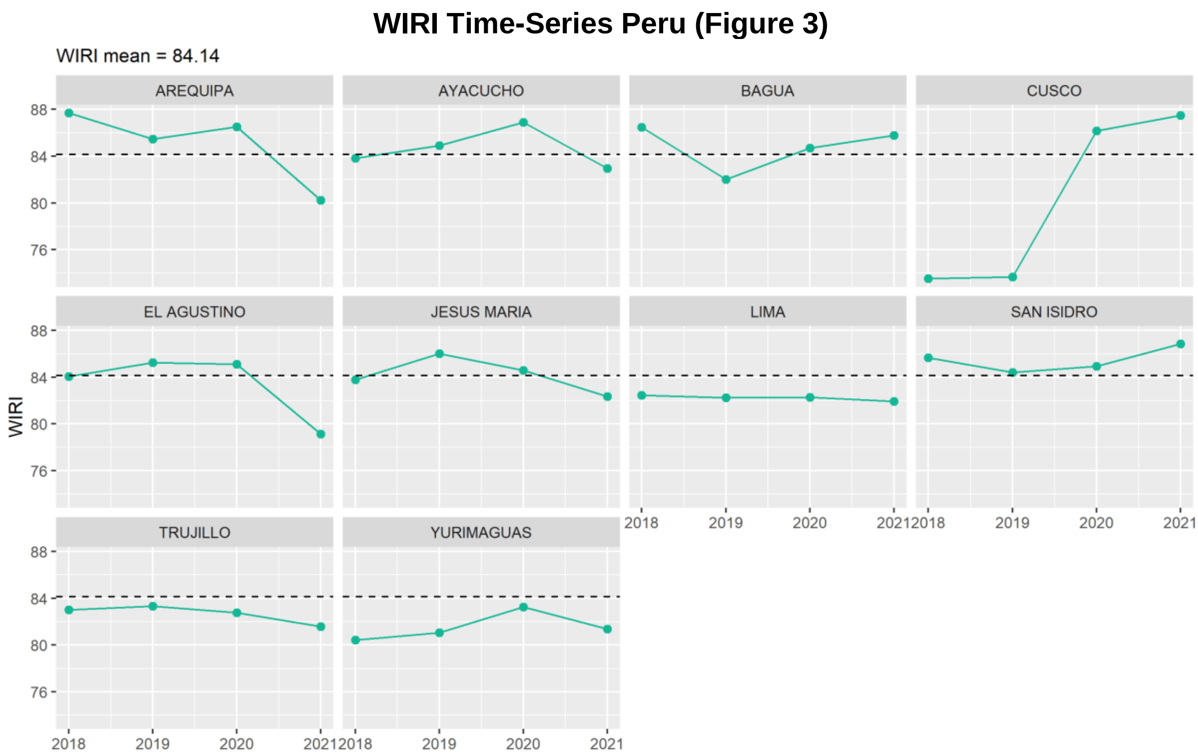
In the case of Bangladesh, Chandpur and Noakhali score considerably lower than their peers, though only Chandpur has a penalized pillar (interactions). The largest cities – Dhaka and Chattogram – constitute the bulk of the observations at the procurement level. Nevertheless, there is a clear distinction between them, with Dhaka being on the lower end of the WIRI spectrum due in part to its low score in the interactions pillar. In 7 of the 10 urban areas, the Operations pillar is the main cause that reduces the average cross-sectional WIRI score.

For Kenya, three additional cities were added. These urban areas were selected based on the availability of W&S contracts. Nyeri and Siaya score considerably lower than Mombasa and Nairobi mostly due to the penalization of missing data on client utility interactions. In the case of Nyeri however, this is slightly compensated by a high integrity score on the operations pillar. Mombasa has the highest levels of integrity in Kenya according to WIRI, followed by the capital. It is important to note, however, that the majority of procurement-level-W&S observations are concentrated in Nairobi. We considered 5 urban areas for this report (one later being dropped) with the highest number of W&S procurement observations, however approximately 70% of those were in the capital.

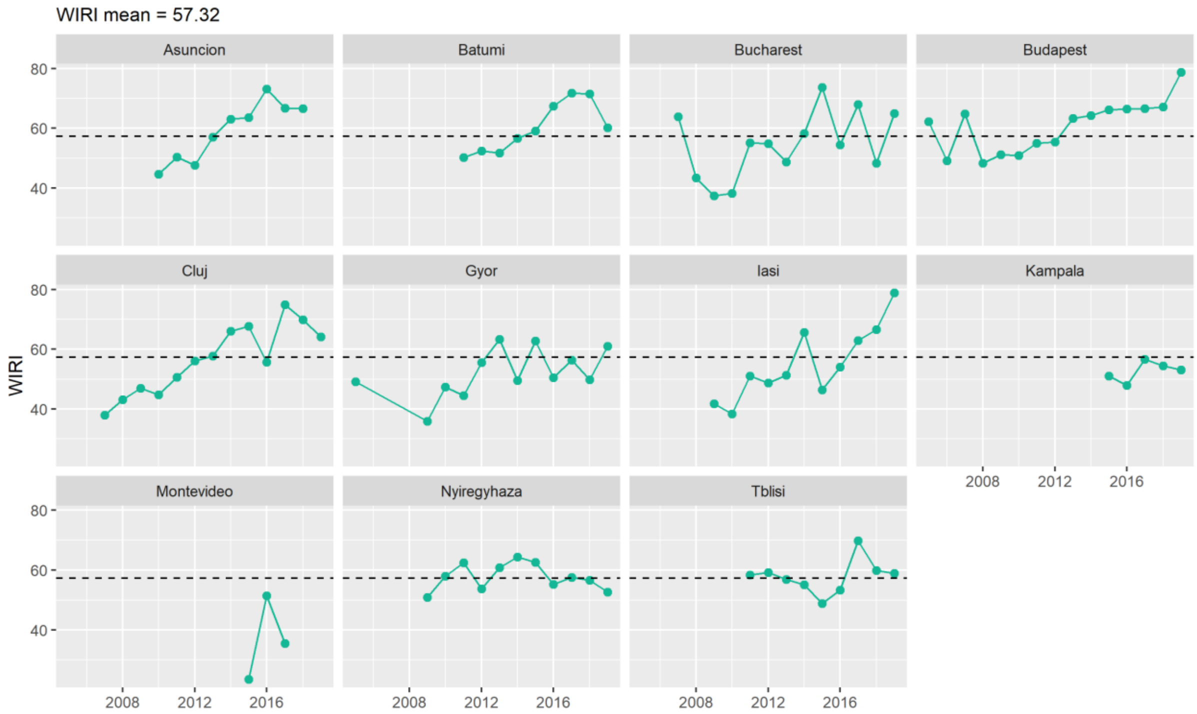
Finally, in the case of Peru, we can observe less variation both in the index scores for the analyzed urban areas, as well as across the individual pillars of WIRI. Nevertheless, scores on the Interactions pillar tend to be slightly lower, and to display slightly larger variation. Not only do we observe higher values for Peruvian urban areas overall, but we also have fewer cases of missing data. Cusco and San Isidro lead the chart, though scores and their components tend to be clustered around the mean. It is important to note, that we currently do not have survey data on direct experiences with corruption in the W&S sector included in the calculations (see Table 3).

Next, we present the dynamics of the WIRI index over time. Figures 3-6 shows the evolution of WIRI scores per urban area whenever we had sufficient data (at least 5 contracts per year per pillar). One of the main challenges we face in this calculate stems from the fact that a significant portion of the variation over time in these localities is partly explained by the penalization of an absent pillar as a 0 on a year-by-year basis.

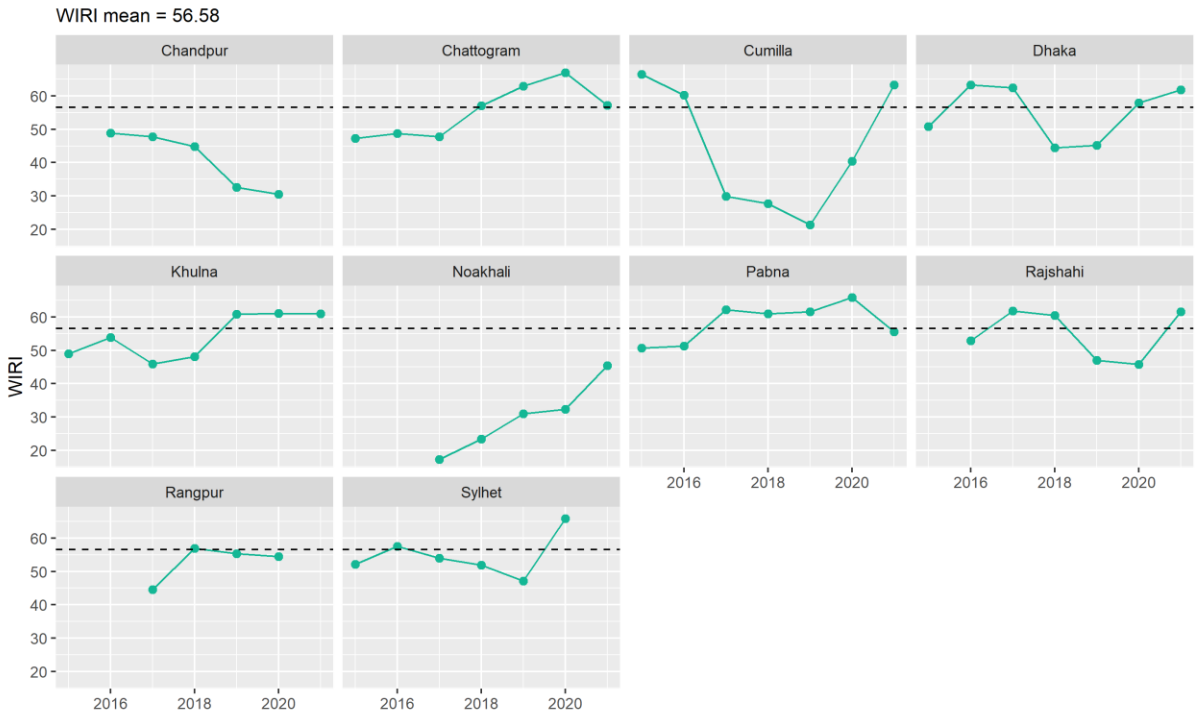
FIGURES 3-6. LONGITUDINAL WIRI BY URBAN AREA (2012-2021)



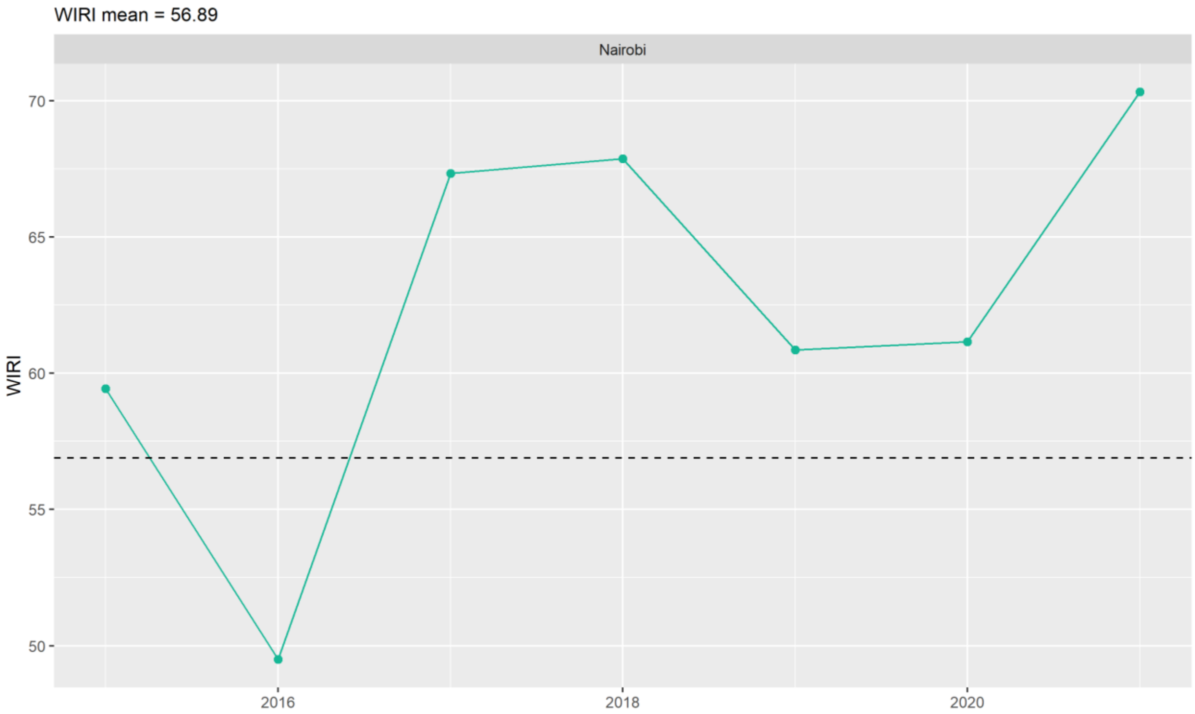
WIRI Time-Series Pilot (Figure 4)



WIRI Time-Series Bangladesh (Figure 5)



WIRI Time-Series Kenya (Figure 6)



In sum, the Water Integrity Risk Index presented in this paper aims to provide an objective measure of corruption risks in the urban water and sanitation sector. As shown in the preceding sections, WIRI is a replicable, transparent, and scalable index, which enables us to compare risk levels in the sector across time and between cities. Moreover, using WIRI we can also observe variations in the three pillars, and retrieve detailed information about individual indicators that increase or reduce the overall rating of the index. These attributes of WIRI makes it a potentially useful measurement for all stakeholders, especially policy makers in the W&S sector.

There are some limitations inherent to the index methodology presented throughout this paper. Namely, the lack of availability of data – especially when it comes to surveys – presents significant challenges. We attempt to circumvent this issue by weighing each pillar based on the global availability of data on its components. Thus, indicators that have high global data sparsity will carry less weight and, consequently, the absence of a datapoint in a urban area when that data is widely available for other urban areas decreases its overall WIRI score. Furthermore, this approach allows for greater flexibility for future iterations of the index as better data becomes more available. Despite such limitations, the WIRI is a robust and replicable measure of corruption in the W&S sector that is based on objective data and thus less prone to the biases characteristic of measuring perceptions of corruption.

5. Lessons Learned

The main methodological update we have introduced to this new WIRI iteration has been the issue of penalization. In contrast to the 2020 pilot version, we recommend only calculating the time-series WIRI score for urban areas where we have data on at least two pillars in any given year. This ensures that we avoid distortions in the interpretation of the aggregate score. Because of this adjustment, the current methodology penalizes missing values in cases where only one of the three pillars is missing. This falls under the general index design logic of taking commonly reported yet missing data as a red flag.

One issue that remains is that of benchmarking and cross-national comparisons. Appendix D shows quartile distribution of WIRI scores and their components for all urban areas using the weights from the pilot study across the board. Furthermore, closer cooperation with local partners should improve the index's granularity by a) providing more data on public procurement contracts that may not be centrally reported, b) expanding the scope of survey data and pipe lengths at the level of urban areas, and c) providing feedback during the water contract classification stage to ensure that all relevant observations and actors are identified.

Finally, we have applied the same weights to each pillar and indicator to the new urban areas based on the first iteration of the WIRI in 2019-20. The original weights were determined based on the overall availability of data for countries in the pilot study. Moving forward, we should recalculate these weights to consider the new countries and settlements added to WIRI or consider a different weight strategy such as country-specific weights.

Bibliography

Andrés, L. A., Schwartz, J., & Guasch, J. L. (2013). Uncovering the Drivers of Utility Performance. In Uncovering the Drivers of Utility Performance. The World Bank. https://doi.org/10.1596/978-0-8213-9660-5

Atangana Ondoa, H. (2017). The effects of Heavily Indebted Poor Countries Initiative (HIPC) on access to safe drinking water. International Review of Public Administration, 22(2), 107–122. https://doi.org/10.1080/12294659.2017.1316947

Auriol, E., & Blanc, A. (2009). Capture and corruption in public utilities: The cases of water and electricity in Sub-Saharan Africa. Utilities Policy, 17(2), 203–216. https://doi.org/10.1016/j.jup.2008.07.005

Buehn, A., & Schneider, F. (2012). Corruption and the shadow economy: Like oil and vinegar, like water and fire? International Tax and Public Finance, 19(1), 172–194. https://doi.org/10.1007/s10797-011-9175-y

Baillat, a. (2013). Corruption and the human right to water and sanitation: Human right-based approach to tackling corruption in the water sector. In Corruption and the human right to water and sanitation. http://www.waterlex.org/new/wp-content/uploads/2013/12/2013-WaterLex-WIN_Corruption-and-the-HRWS-.pdf

Bardhan, P. (1997). Corruption and Development: A Review of Issues. Journal of Economic Literature, 35(3), 1320–1346. https://doi.org/10.4324/9781315126647-30

Chakraborty, A., Ghosh, S., Mukhopadhyay, P., Dinara, S. M., Bag, A., Mahata, M. K., Kumar, R., Das, S., Sanjay, J., Majumdar, S., & Biswas, D. (2014). Trapping effect analysis of AlGaN/InGaN/GaN Heterostructure by conductance frequency measurement. In MRS Proceedings: Vol. XXXIII (Issue 2). World Bank Publications. https://doi.org/10.1007/s13398-014-0173-7.2

Das, B., der Gaag, N., Rychlewsk, M., McIntyre, P., & Fernández, C. F. (n.d.). Water integrity Global Outlook.

Davis, J. (2004). Corruption in public service delivery: Experience from South Asia's water and sanitation sector. World Development, 32(1), 53–71. https://doi.org/10.1016/j.worlddev.2003.07.003

Dillon, W. R., Bollen, K. A., & Long, J. S. (1996). Testing Structural Equation Models. In Journal of Marketing Research (Vol. 33, Issue 3). https://doi.org/10.2307/3152134

Earle, A. (2007). The role of governance in countering corruption: An African case study. Water Policy, 9(SUPPL. 2), 69–81. https://doi.org/10.2166/wp.2007.131

Fazekas, M., & Kocsis, G. (2020). Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data. British Journal of Political Science, 50(1), 155–164. https://doi.org/10.1017/S0007123417000461

Fazekas, M., Tóth, I. J., & King, L. P. (2016). An Objective Corruption Risk Index Using Public Procurement Data. European Journal on Criminal Policy and Research, 22(3), 369–397. https://doi.org/10.1007/s10610-016-9308-z

Fazekas, M., et. al. (2020). Beyond leakages Quantifying the Effects of Corruption on the Water and Sanitation Sector in Latin America and the Caribbean, IADB Working Paper, https://publications.iadb.org/publications/english/document/Beyond-Leakages-Quantifying-the-Effects-of-Corruption-on-the-Water-and-Sanitation-Sector-in-Latin-America-and-the-Caribbean.pdf

Gomez, M., Perdiguero, J., & Sanz, àlex. (2019). Socioeconomic factors affecting water access in rural areas of low and middle income countries. Water (Switzerland), 11(2), 202. https://doi.org/10.3390/w11020202

Gulati, M., & R a o, M. . (2007). Corruption in the Electricity Sector: A Pervasive Scourge. In The many faces of corruption. Tracking vulnerabilities at the sector level (pp. 114–157). World Bank.

Hall, D., & Lobina, E. (2007). International actors and multinational water company strategies in Europe, 1990-2003. Utilities Policy, 15(2), 64–77. https://doi.org/10.1016/j.jup.2007.02.005

Halpern, J., Kenny, C., Dickson, E., Ehrhadrt, D., & Chloe, O. (2018). Deterring Corruption and Improving Governance in the Urban Water Supply & Sanitation Sector. A Sourcebook. Water Working Notes, 18. http://documents.worldbank.org/curated/en/119321468140052572/pdf/442240REPLACEM1ument10WN141TownsWSS.pdf

Hoyle, R. (2012). Handbook of structural equation.

Jergelind, V. (2015). Facing the challenges of water efficiency. In World Pumps (Vol. 2015, Issue 3). https://doi.org/10.1016/S0262-1762(15)30028-6

Kirkpatrick, C., Parker, D., & Zhang, Y. F. (2006). An empirical analysis of state and private-sector provision of water services in Africa. World Bank Economic Review, 20(1), 143–163. https://doi.org/10.1093/wber/lhj001

Klašnja, M. (2017). Uninformed Voters and Corrupt Politicians. American Politics Research, 45(2), 256–279. https://doi.org/10.1177/1532673X16684574

Knack, S., & Keefer, P. (1995). Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures. Economics & Politics, 7(3), 207–227. https://doi.org/10.1111/j.1468-0343.1995.tb00111.x

Lemke, C. (2008). On the benefit of using time series features for choosing a forecasting method. 2nd European Symposium on Time Series Prediction. http://eprints.bournemouth.ac.uk/8514/

Lütkepohl, H. (2004). Univariate time series analysis. In Applied Time Series Econometrics. https://doi.org/10.1017/CBO9780511606885.003

Molenberghs, G., & Verbeke, G. (2013). Missing data. In The SAGE Handbook of Multilevel Modeling. London: SAGE Publications Ltd. https://doi.org/10.4135/9781446247600.n23

Sarafidis, V., & Wansbeek, T. (2012). Cross-Sectional Dependence in Panel Data Analysis. Econometric Reviews, 31(5), 483–531. https://doi.org/10.1080/07474938.2011.611458

Levy, B. (n.d.). Governance Reform: Bridging Monitoring and Action. World Bank Publications.

Makoni, P. L. (2014). Infrastructure quality, firm characteristics and corruption in Tanzania. Corporate Ownership and Control, 12(1 Continued 4), 379–385. https://doi.org/10.22495/cocv12i1c4p2

Marek Chromý, Milan Eibl, Nemanja Nenadic, Zlatko Minic, Z. D. (2019). SOEs and MOEs Transparency index (SAMET Index - PETRA). https://www.transparency.cz/wp-content/uploads/2020/01/SOEs-and-MOEs-Transparency-index.pdf

Michael, J. (1996). The Search for Definitions: The Vitality of Politics and The Issue of Corruption. International Social Science Journal, 48(149), 321.

Mungiu-Pippidi, A. (2006). Corruption: Diagnosis and treatment.(political corruption). Journal of Democracy, 17(3), 86.

Mungiu-Pippidi, A. (2014). The Anticorruption Frontline - The Anticorruption Report. ProtoView, 2(47). http://search.proquest.com/docview/1651914489/

Nauges, C., & van Den Berg, C. (2010). Heterogeneity in the cost structure of water and sanitation services: A cross-country comparison of conditions for scale economies. Oxford Development Studies, 38(2), 199–217. https://doi.org/10.1080/13600811003753768

North, D. C., Wallis, J. J., & Weingast, B. R. (2009). Violence and social orders: A conceptual framework for interpreting recorded human history. In Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History. Cambridge University Press. https://doi.org/10.1017/CBO9780511575839

Nye, J. S. (1967). Corruption and Political Development: A Cost-Benefit Analysis. American Political Science Review, 61(2), 417–427. https://doi.org/10.2307/1953254

OECD. (2007). Bribery in public procurement: Methods, actors and counter-measures. In Bribery in Public Procurement: Methods, Actors and Counter-Measures (Vol. 9789264013). OECD. https://doi.org/10.1787/9789264013964-en

Pituch, K. A. (2015). Applied Multivariate Statistics for the Social Sciences. In Applied Multivariate Statistics for the Social Sciences. https://doi.org/10.4324/9781315814919

Punjabi, B. (2017). Canal bureaucracy and the corruption nexus around water in the Mumbai hinterland: Questions for development and governance in Maharashtra, India. India Review, 16(2), 179–211. https://doi.org/10.1080/14736489.2017.1314137

Pusok, K. (2016). Public-Private Partnerships and Corruption in the Water and Sanitation Sectors in Developing Countries. Political Research Quarterly, 69(4), 678–691. https://doi.org/10.1177/1065912916658552

Rafi, M. M., Lodi, S. H., & Hasan, N. M. (2012). Corruption in Public Infrastructure Service and Delivery: The Karachi Case Study. Public Works Management and Policy, 17(4), 370–387. https://doi.org/10.1177/1087724X12450642

Rothstein, B., & Teorell, J. (2008). What is quality of government? A theory of impartial government institutions. Governance, 21(2), 165–190. https://doi.org/10.1111/j.1468-0491.2008.00391.x

Sarafidis, V., & Wansbeek, T. (2012). Cross-Sectional Dependence in Panel Data Analysis. Econometric Reviews, 31(5), 483–531. https://doi.org/10.1080/07474938.2011.611458

Shleifer, A., & Vishny, R. W. (1993). Corruption. The Quarterly Journal of Economics, 108(3), 599–617. http://www.jstor.org/stable/2118402

Appendix A. Reviewed Sources for Corruption In W&S

Source name (survey/stats office/utility website)	Fee paying subscription source	Local data availability	Reason why it isn't relevant
Aquarating	No	No	evaluation of the utilities' performance no public data.
Bribe Payers index 1999	No	No	It covers general corruption, not water sector corruption.
Bribe Payers index 2006	No	No	It covers general corruption, not water sector corruption.
Bribe Payers index 2008	No	No	It covers general corruption, not water sector corruption.
Bribe Payers index 2011	No	No	It covers general corruption, not water sector corruption.
Chile National Statistical Office	No	No	no structured data
Colombia National Statistical Office	No	No	no structured data
Corruption perception index 1995 - 2018	No	No	Asks business executives about perception of corruption, not direct corruption.
Ecuador National Statistical Office	No	No	no structured data
EPMAPS - Quito utility company	No	No	no structured data
ESSAP utility company	No	Some	no structured data
Eurobarometer 2011	No	No	Asks about perception of bribing, not bribing itself. Water services are not included.
European Quality of Government Index (EQI) 2010-2013	No	No	Questions on corruption in health services, police and government run agencies.
European Quality of Government Index (EQI) 2017	No	No	Questions on corruption in health services, police and government run agencies.
European social survey 2002- 2003	No	No	Asks about bribery in general, not water services specific.
European Social Survey 2004	No	No	Asks about bribery in general, not water services specific.
European social survey 2005	No	No	Asks about bribery in general, not water services specific.
European social survey 2018	No	No	Asks about bribery in general, not water services specific.
European values study 1981 - 1990	No	No	No questions on bribery in water services.
European Values Study 1999	No	No	Asks about individual citizen values and justification of corruption. Water services not included.
European Values Study 2008	No	No	Asks about individual citizen values and justification of corruption. Water services not included.

GEMI	No	No	Wastewater treatment and water quality no local or public structured data
GLAAS WHO	No	No	The survey contains questions around water policy and indicators of legal aspects of access etc, but the results are country level, not on local level data \LINK
Global Corruption Barometer 2003	No	No	It asks people about their perception around corruption.
Global Corruption Barometer 2015-2017	No	No	It covers general corruption, not water sector corruption.
Global Corruption Barometer Transparency International 2003	Yes	No	It is only country level results in percentages data there is no microdata
Global Corruption Barometer Transparency International 2005	Yes	No	It is only country level results in percentages data there is no microdata
Global Corruption Barometer Transparency International 2009	Yes	No	It is only country level results in percentages data there is no microdata
http://waterintegritynetwork.net/?s=%22survey%22	No	No	No availability of microdata
https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NRMRL&dirEntryId=200508	No	No	No availability of microdata
https://www.water.org.uk/wp-content/uploads/2018/12/GWI-International-sector-performance-comparisons.pdf	No		No availability of microdata
IBNet data	No	Some	some, it differs on sample size per country and the years may not match.
Interagua utility company	No	No	no structured data
International Country Risk Guide assessment political risk service	No	No	Private generated reports on corruptions risks per country.
Jamaica National Statistical Office	No	No	no structured data
Latinobarometer 1995 - 2018	No	No	Asks citizens about quality of water services after privatization.
Mexico City Procurement Data	No	No	The project is ongoing and the official source of data is unavailable
Mexico National Statistical Office	No	Some	no structured data
OSE utility company	No	No	no structured data
Paraguay National Statistical Office	No	Some	no structured data
Peru National Statistical Office	No	No	no structured data
Political Risk Service - Corruption by the The PRS Group.	Yes	No	It is not water sector specific.
SDG 6 2016 data on Water and Sanitation	No	No	It is only country level data indicators for quality and access.

The Quality of Governance Expert Survey 2015	No	No	Applied only to business executives about corruption in general from public administration, divided into powers
UN resources	No	No	Available data is country level data, and it is produced on the basis of statistics offices of countries.
United Nations Development Data 2000, 2005, 2010,2015	No	No	Available data is country level data, and it is produced on the basis of statistics offices of countries.
Uruguay National Statistical Office	No	Some	no structured data
WaCClim Climate Smart Water	No	No	Tool kits, reports, assessments and case studies on water quality and governance
WASH Joint Monitoring Programme for Water Supply, Sanitation and Hygiene 2000 to 2017	No	No	Data collected from national statistics offices and aggregated by country into country reports. No microdata.
Water funds toolbox	No	No	Tool kits, reports, assessments and case studies on water.
Water safety portal	No	No	Tool kits, reports, assessments and case studies on water.
World bank development data	No	No	It is only country level data indicators for quality and access.
World Values Surveys (Wave 6, 2010-2014)	No	Yes	It covers general corruption, not water sector corruption.

wi

Appendix B. Keywords for searches in public procurement data

General categories of terms in English include: the specific water utility name to each urban area.

- Paraguay. Asunción . Terms include: “servicios + sanitarios”, “Alberdi”, “San Bernardino”, “Erssan”, “sistema & agua”, “saneamiento & agua”, “Cuenca & agua”, “alcantarillado + agua”, “drenaje & agua”, “servicio red + agua”, “Saneamiento & conexiones”, “constr & agua”, “Cuenca”, “tuberias & agua”, “canal & agua”, “sistemas de abastecimiento & agua”, “empresa & servicios & sanitarios & Paraguay”, “gerencia & redes & Asunción & metropolitana”.
- Peru. All. Terms include: “servicios + sanitarios”, “sistema & agua”, “saneamiento & agua”, “Cuenca & agua”, “alcantarillado + agua”, “drenaje & agua”, “servicio red + agua”, “Saneamiento & conexiones”, “constr & agua”, “Cuenca”, “tuberias & agua”, “canal & agua”, “sistemas de abastecimiento & agua”, “empresa & servicios & sanitarios & Peru”, “gerencia & redes & metropolitana”.
- Hungary. Budapest, Győr, Nyíregyháza. Terms include: “vizikozmu”, “vizugyi”, “szennyviztisztito”, “Vizmuvek”, “vizikozmu szovetseg”, “Kozuzemi”, “szennyviz”, “ivoviz minoseg”, “csapadek viz”, “Szennyvizcsatorn”, “pannon, nyirsegviz”.
- Romania. Cluj, Bucharest, Iasi. Terms include: "apa + violia" "anrsc "name of utility"apa + nova"apa + uzata "apa + uzata + glina"apa + utilitatea"salubritate"sanitatiaon "sanitar"sanitary "canalizare"sewerage"se distileaza"distill "apa + canal" watercanal "apa + retea" water network "apa + constructie" water construction "apa + constructia" water construction "apa + teava" water pipe "apa + livra" water supply"apa + rezerva" water supply/reserve"apa + sistem" water system "apa + testarea" water testing" apa + distilare "water distill "apa + functioneaza "water works "apa + uzina" water works" apa + reziduale" waste water" apa + lucrari" water works
- Georgia. Batumi, Tbilisi. Terms include: utility name in Georgian. Georgian Water and Power (GWP),
- Uganda. Kampala. Terms include: ministry + water , national water , sanitation , sanitary , sewer , water + network , water + construction , water + channel , water + pipe , water + sewerage , water + supply + drinking , water + system , water + testing , water + construction , water + district , water + distill , water + works , national + water + sewerage , kampala + water , pipe , pipeline
- Bangladesh. All. Terms include: ministry + water , national water , sanitation , sanitary , sewer , water + network , water + construction , water + channel , water + pipe , water + sewerage , water + supply + drinking , water + system , water + testing , water + construction , water + district , water + distill , water + works , national + water + sewerage , pipe , pipeline
- Kenya. Nairobi. Terms include: ministry + water, national water, nairobi+metropolis, athi+water+works, wasre, water+sanitation, irrigation, housing+development, water+project, nairobi+sanitation, water+authority, nairobi+sewerage, kenya+water, kenya+water+towers, kenya+water+institute, nationa+water, pipeline+water, sewer, sanitation, sanitary, sewer, water + network, water + construction, water + channel, water + pipe, water + sewerage, water + supply + drinking, water + system, water + testing, water + construction, water + district, water

+ distill, water+treatment, water+pipeline, pipeline+extension, water+desilting, water+guttering, water+rain+collection, water + works, national + water + sewerage, nairobi + water, water + athi, pipe, pipeline

Appendix C. Variable Dictionary

Variable name	Definition	Integrity Pillar	Source
avg_cri_inv	Calculation of CRI for investment = (Average single biding Indicator aggregation of the single bidding component + average Length of decision period investment indicator + length of investment period + No call for tenders publication indicator + Procedure type)/5	Investment	Calculations produced from Public Procurement Data
avg_cri_inv_100	Average corruption indicators normalized	Investment	Calculations produced from Public Procurement Data
count_inv	Total number of investment contracts	Investment	Calculations produced from Public Procurement Data
contract_value_inv	Total value of investment contracts in local currency	Investment	Calculations produced from Public Procurement Data
pipe_investment_value_local	total value of pipe investment contracts in local currency	Investment	Calculations produced from Public Procurement Data
pipe_contracts_count	Number of pipe contracts	Investment	Calculations produced from Public Procurement Data
pipelength	Length of network of pipes in Km	Investment	Statistical offices of countries
total_pipe_valueinUSD	Value of pipe contracts in international USD pipe_investment_value_local / bf_wb_ppp	Investment	Calculation
yhat	prediction of pipe investment using regression analysis	Investment	Calculation
resid	residuals of pipe investment using regression analysis	Investment	Calculation
pipe_int	Pipe investment indicator normalized residuals *100	Investment	Calculation
wiri_inv	Investment WIRI indicator, average corruption indicator normalized and pipe investment indicator	Investment	Calculation
minresid	Minimum residuals, standardized	Investment	Calculation
maxresid	maximum residuals from pipe investment standardized	Investment	Calculation
wiri_ops	WIRI operations indicator average of operation integrity indicator	Operations	Calculation

Variable name	Definition	Integrity Pillar	Source
avg_cri_op	Calculation of CRI for operations = (Average single biding Indicator aggregation of the single bidding component + average Length of decision period investment indicator + length of investment period + No call for tenders publication indicator + Procedure type)/5	Operations	Calculations produced from Public Procurement Data
avg_cri_op_int_100	Average corruption indicators normalized	Operations	Calculations produced from Public Procurement Data
count_op	Total value of operations contracts in local currency	Operations	Calculations produced from Public Procurement Data
contract_value_opsIUSD	Value of contracts in operations transformed into international USD (contract_value_main/ bf_wb_ppp)	Operation	Calculation
water_settlement	Coded water settelements 1- Asuncion, 2 - Batumi, 3 - Bucharest, 4 - Budapest, 5 - Cluj, 6 - Gyor, 7- Iasi , 8- Kampala, 9- Montenegro, 10- Nairobi, 11 - Nyíregyháza, 12 -Tibisili	Identifying information	Unique
tender_year	year of tenders	Identifying information	Public Procurement Data
avg_cri_inter	Calculation of CRI for client utility interaction = (Average single biding Indicator aggregation of the single bidding component + average Length of decision period investment indicator + length of investment period + No call for tenders publication indicator + Procedure type)/5	Client Utility Interaction	Calculations produced from Public Procurement Data
Pipel_int	Infrastructure interaction from regression model, missing infrastructure.	Investment	Calculations
contract_value_inv_IUSD	value of total investment contracts in international USD (contract_value_inv/ bf_wb_ppp)	Investment	Calculation
avg_cri_inter_int_100	Average corruption indicators normalized	Client Utility Interaction	Calculations produced from Public Procurement Data
contract_value_inter	Total value of client utility contracts in local currency	Client Utility Interaction	Calculations produced from Public Procurement Data

Variable name	Definition	Integrity Pillar	Source
cui_beeps_bribery	number of BEEP respondents to yes on bribery out of sample size	Client Utility Interaction	BEEPS
cui_beeps_bribery_int	cui_beeps_bribery_int = (cui_beeps_bribery_total - 0)/(cui_samplesize_beeps- 0)*(100)-100	Client Utility Interaction	Calculation
cui_afrb_bribery	number of Afrobarometer respondents to yes on bribery out of sample size	Client Utility Interaction	Afrobarometer (data requested through email)
cui_afrb_bribery_int	cui_afrb_bribery_int = (cui_afrobarometer_total - 0)/(cui_samplesize_afrobarometer -0)*(100)-100	Client Utility Interaction	Calculation
cui_beeps__bribery_total	Number of respondents that admitted to bribery in the water sector	Client Utility Interaction	BEEPS
wiri	WIRI indicator (wiri_inv + wiri_ops + wiri_inter)/3	Client Utility Interaction	Calculation
cui_samplesize_beeps	Total number of sample size of BEEPS survey in each urban_area	Client Utility Interaction	BEEPS
cui_afro_barometer_total	Number of respondents that admitted to bribery in the water sector	Client Utility Interaction	Afrobarometer (data requested through email)
cui_samplesize_afrobaromer	Total number of sample size of Afrobarometer in each settlement	Client Utility Interaction	Afrobarometer (data requested through email)
contract_value_inter_IUSD	Total value of contracts in the client utility interaction (contract_value_inter/ bf_wb_ppp)	Client Utility Interaction	Calculations produced from Public Procurement Data
contract_value_main	Total value of operation contracts in local currency	3 Pillars	Calculations produced from Public Procurement Data
contract_value_total_local	Total value of combined contracts (investment, operations, client utility interactions) in local currency	3 Pillars	Calculations produced from Public Procurement Data
count_total	Total number of contracts	3 Pillars	Calculations produced from Public Procurement Data
contract_value_total_IUSD	Value of total contracts in international USD	3 Pillars	Calculation
bf_wb_ppp	Price parody controlled for inflation	3 Pillars	World Bank
wiri_inter	WIRI client utility interaction Integrity Indicator	3 Pillars	Calculation

Variable name	Definition	Integrity Pillar	Source
	(avg_cri_inter_int_100 + cui_beeps_bribery_int + cui_afrb_bribery_int)/3		

Appendix D. Quartile Benchmarks

Urban Area	Country	WIRI	wiri_inv	wiri_ops	wiri_cui	WIRI_Q4	inv_Q4	ops_Q4	cui_Q4
Asuncion	Paraguay	68.61	71.21	68.26	61.97	3	3	3	2
Batumi	Georgia	59.57	60.69	58.23	59.15	2	2	2	2
Bucharest	Romania	56.35	46.66	75.87	44.29	2	1	3	1
Budapest	Hungary	61.66	47.93	73.18	77.24	3	1	3	3
Cluj	Romania	50.47	49.76	76.41	0	1	1	3	1
Gyor	Hungary	59.81	48.19	71.03	70	2	1	3	3
Iasi	Romania	50.81	48.64	78.95	0	1	1	3	1
Kampala	Uganda	52.62	52.82	53.06	51.16	1	1	1	2
Montevideo	Uruguay	49.76	41.33	42.18	89.02	1	1	1	4
Nyiregyhaza	Hungary	58.73	48.44	68.05	69.02	2	1	2	2
Tblisi	Georgia	62.08	65.92	62.67	50	3	3	2	1
Chandpur	Bangladesh	41.63	64.13	30.69	0	1	2	1	1
Chattogram	Bangladesh	61.58	66.63	49.71	71.3	3	3	1	3
Cumilla	Bangladesh	51.34	62.08	32.85	58.33	1	2	1	2
Dhaka	Bangladesh	54.63	60.74	55.16	36.23	2	2	1	1
Khulna	Bangladesh	61.24	63.53	50	77.5	2	2	1	3
Noakhali	Bangladesh	43.73	57.81	14.76	62.5	1	1	1	2
Pabna	Bangladesh	67.12	67.35	62.91	75	3	3	2	3
Rajshahi	Bangladesh	60.72	62.72	59.8	56.94	2	2	2	2
Rangpur	Bangladesh	59.91	72.3	45.91	53.13	2	3	1	2
Sylhet	Bangladesh	63.92	66.71	58.66	66.67	3	3	2	2
Mombasa	Kenya	65.36	65.17	63.33	70	3	3	2	3
Nairobi	Kenya	56.98	63.57	59.37	33.48	2	2	2	1
Nyeri	Kenya	53.32	60.22	70	0	1	2	3	1
Siaya	Kenya	51.89	64.42	60	0	1	2	2	1
Arequipa	Peru	84.97	84.66	87.15	81.46	4	4	4	4
Ayacucho	Peru	84.64	84.69	84.65	84.47	4	4	4	4
Bagua	Peru	84.74	84.16	84.01	87.87	4	4	4	4
Cusco	Peru	87.56	88.07	87.55	86.17	4	4	4	4
El Agustino	Peru	83.39	82.75	85.26	81.41	4	4	4	4
Jesus Maria	Peru	84.18	84.03	84.57	83.81	4	4	4	4
Lima	Peru	82.24	81.69	83.48	81.27	3	3	3	4
San Isidro	Peru	85.47	87.13	85.29	81.14	4	4	4	3
Trujillo	Peru	82.67	82.3	84.98	79.06	4	4	4	3
Yurimaguas	Peru	81.53	81.23	83.04	79.33	3	3	3	3