

Citywide surveys of water and sanitation service levels: design and methodology

Urban Sanitation
Research Initiative



Contents

1. Introduction	4
1.1 WSUP's citywide survey aims	4
1.2 Fundamental survey design decisions	5
2. Sampling design	7
2.1 What is your overarching aim?	7
2.2 What is the available budget?	7
2.3 What sample size can you afford?	8
2.4 Is that sample size sufficient to achieve the confidence and precision you need?	8
2.5 How many Enumeration Areas can you achieve?	8
2.6 Identifying Enumeration Areas	9
2.7 Dividing Enumeration Areas into PSUs	10
2.8 Household selection protocol	11
3. Data collection tools and process	12
3.1 Questionnaire	12
3.2 Additional data collection for assessment of safely managed services	13
3.3 Survey delivery	13
4. Quality assurance and data cleaning	14
4.1 Quality assurance	14
4.2 Data cleaning	14
5. Data analysis	15
5.1 Indicators and indicator mapping	15
5.2 Weighting by population size	15
5.3 Data analysis summary	16
6. Survey reporting and visualisation	17
7. Systematic spatial sampling versus other sampling approaches	18
8. Method limitations	20
9. Concluding remarks	22

1. Introduction

This publication describes the methodology used by Water & Sanitation for the Urban Poor (WSUP) in citywide surveys of water and sanitation service levels in 7 cities in 2017/2018. These were surveys designed to give an assessment of water and sanitation service levels across the city, with a particular focus on low-income areas. Surveys were relatively low-cost (budget per city \$10,000 - \$25,000), with total sample size in each city ranging from 600 to 2400 households. We used a type of *systematic spatial sampling*, which we explain further in the text. Design and delivery of surveys of this type requires multiple decisions, and we have produced this publication in the view that this recent WSUP experience may be useful to other organisations who wish to carry out low-cost citywide surveys of WASH service levels. We present our experience as adaptable guidance, not as an inflexible recipe: each survey will of course have different aims and constraints

1.1 WSUP's citywide survey aims

WSUP's citywide surveys carried out 2017/2018 aimed to provide a good understanding of water and sanitation service levels across each of 7 cities (Dhaka in Bangladesh; Maputo in Mozambique; Nakuru in Kenya; Antananarivo, Mahajanga and Toliara in Madagascar; and Lusaka in Zambia). In particular, we wanted reasonably accurate estimates of the proportion of people in each of the "service ladder" categories for water and sanitation as defined by the WHO/UNICEF Joint Monitoring Programme (JMP). Our primary interest was in WASH service levels in low-income communities. We had budget of up to \$25,000 available for each survey, and a team of 2 people in London with strong time dedication to survey design and management.

These surveys were carried out primarily as baseline surveys within our Business Plan 2016–2020, to be paralleled by equivalent endline surveys in 2019/2020, allowing assessment of progress toward universal WASH coverage in these cities in which we work.¹ This tracking approach (as opposed to simple tracking of progress in specific areas of the city in which WSUP has directly supported service delivery) is aligned with our Theory of Change, which focuses on system change at the city and national levels; but certainly it raises attribution

questions, because WSUP is of course not the only actor working to improve WASH service levels in these cities. Furthermore, it raises questions around likely rate of progress towards universal WASH coverage: it's unlikely that major citywide progress will be made within the period 2017–2020, so our endline surveys are likely to be smaller and more focused, with likely repeat of the more extensive surveys within our next Business Plan 2021–2025.²

Although primarily delivered to meet our own organisational needs for progress tracking, we of course wanted these surveys to be useful to other actors, including but not limited to water and sanitation utilities, municipal governments, national government agencies, other development actors working in those cities, and international progress-monitoring initiatives such as the JMP³ and GLAAS. Indeed, generating more useful survey data was a core driver of our decision to deliver citywide surveys: typical NGO baseline/endline surveys in specific intervention districts are of very limited value to other actors, whereas periodic citywide surveys are potentially of much greater value.

Local actors made clear to us that these surveys would be more useful if we could achieve some degree of disaggregation by administrative boundaries, allowing service levels in different major sub-divisions of the city to be compared.

¹ These are not the only cities or countries in which we work: but they're all major cities in which we have had a significant long-term presence, so we chose these 7 cities as representative of our work for M&E (monitoring and evaluation) purposes and based on WSUP in-country management view of where this data would be most useful.

² We note here that tracking of change over time is more precise if the same sampling units are subsequently revisited.

³ One component of this survey process has been to identify workable ways of assessing aspects required for consideration of service levels as "safely managed" (notably water quality, and proportion of faecal waste safely treated), and we hope this experience may be useful to the JMP and other actors.

So these were citywide surveys with intra-city spatial disaggregation: the nature of the spatial disaggregation varied from city to city, but in most cases we obtained disaggregated estimates for low-income area groupings within administrative sub-divisions of the city. Surveys had between 4 and 11 sub-divisions: as explained in Section 2.4, each additional sub-division increases total sample size by around 100 households.

As noted, we were primarily interested in low-income areas, in view of WSUP's *raison d'être* as a pro-poor agency: we considered that, in most cities, it is reasonable to assume that water and sanitation service levels in higher-income and commercial areas are more likely to be adequate, so that available sampling resource is more usefully focused on low-income areas.⁴ In particular, high-income areas are less likely to experience major changes in water and sanitation over time (or at any rate, changes are likely to be from good to better, so not centrally relevant for tracking citywide coverage in terms of the JMP ladders);⁵ including them would dilute any estimate of more important changes occurring in low-income areas. In practical terms, our approach allows for specific consideration of a greater number of low-income areas and produces more precise estimates in these, and is therefore more useful for WASH tracking and planning. There is also another practical consideration here: it is often much more difficult to do door-to-door survey in non-poor areas, because many homes form part of gated buildings or condominiums, and it's difficult for enumerators to get past those gates!

This then is a summary of WSUP's aims with these surveys. If you're considering surveys yourself, your aims may of course be similar, or very different.

1.2 Fundamental survey design decisions

In designing WSUP's surveys, as any other survey, we had to make trade-offs between different aims. In our first survey, in Dhaka, we ideally wanted sufficient sampling power to provide:



Image: Collecting water in Kalshi, Dhaka. Credit: GMB Akash

- a) good estimates of WASH service levels across the city (including non-poor areas), disaggregated by administrative division;
- b) good estimates of WASH service levels in low-income communities, again disaggregated by administrative division; and
- c) good estimates of WASH service levels within specific communities in which WSUP had intervened.

Evidently, for every sub-unit of the city for which you want a statistically valid independent estimate, you need a given minimum sample size. Predictably, we found that our total available sample size was insufficient to provide all of **a–c**.

Subsequently, we have simplified our aims, and our surveys focus on **b**, with a small part of the sample dedicated to assessing WASH service levels in non-poor areas for comparison purposes (see comments in Section 1.1 above). Therefore our surveys can essentially be considered citywide surveys of WASH service levels in low-income areas.

In our first 2 surveys, in Dhaka-Bangladesh and Maputo-Mozambique, we used a particular survey methodology developed for WSUP by consultants Valid International. Subsequently we moved to a simpler methodology, which developed over time as we gained a better understanding of what worked best.

⁴ But we stress that this is a relative judgement: non-poor areas may suffer issues such as poor continuity of water supply, and while sanitation will typically be acceptable at the household level, eventual treatment of waste may certainly often be inadequate. For further discussion of this assumption, see Section 8.

⁵ We note that even in high-income areas, sanitation may often be "basic" but not "safely managed", so that our focus on low-income areas is not well-designed for tracking "basic" to "safely managed" transitions in non-poor areas.

Table 1. Summarised design characteristics of the different surveys. The methodology described in the text is that used for the Lusaka survey, which is likely to be the methodology we will use in future.

	DHAKA	MAPUTO	NAKURU	MADAGASCAR	LUSAKA
SAMPLING STRATEGY	Two-stage stratified clustered systematic spatial sampling (slum/non slum areas)	Two-stage clustered systematic spatial sampling	Systematic spatial sampling	Systematic spatial sampling	Systematic spatial sampling
SAMPLE SIZE CALCULATION	Lot Quality Assurance Sampling (with bootstrap analysis) used both for slum areas (N=66) and non-slum areas (N=66) within each EA. For each EA total N = 132 1,282 HH	Adjusted by DEFF Confidence 95% Precision (each EA) $\pm 10\%$	Confidence 95% Precision (each EA) $\pm 10\%$ SS enhanced by 10%	Confidence 95% Precision (each EA) $\pm 10\%$ SS enhanced by 32%	Confidence 95% Precision (each EA) $\pm 10\%$ SS enhanced by 7%
	132 HH per EA (slum = 66 HH, non-slum = 66)	154 HH per EA	110 HH/EA	146 HH per EA	105 HH per EA
	2,376 HH	616 HH	1,210 HH	1,314 (A) + 784 (M) + 730 (T)	1,155 HH
IDENTIFICATION OF PSUs	Map-based spatial systematic sampling (Centric Systematic Area Sampling)				
STUDY POPULATION	Slum and non-slum population	Total population excluding one high-income district.	Low-income urban population, plus 1 middle- and 1 high-income area	Citywide urban population	Low-income urban population plus 1 middle-income area
EA GROUPINGS	9 EAs (18 strata)	4 EAs	11 EAs	9 EAs (A), 6 EAs (M), 5 EAs (T)	11 EAs
	Based on geographical proximity	Based on administrative divisions	Based on administrative divisions, proximity and similarities	Based on administrative divisions, population size, proximity and WSUP intervention	Based on administrative divisions, geographical proximity and service provider

Abbreviations: DEFF = design effect (in cluster sampling). EA = Enumeration Area. HH = households. PSU = Primary Sampling Unit. SS = sample size. Madagascar cities: A = Antananarivo, M = Mahajanga, T = Toliara.

This methodology maintained Valid International's *systematic spatial sampling* methodology (notably their algorithm for gridding survey sub-areas and their protocol for identifying the respondent household within each grid-square), but simplified the objectives, with increased sample allocation to low-income communities and reduced focus on middle/high-income communities. The sampling approach was also revised, moving away from clustered sampling (two households per grid-square) to one household per grid-square.

This increased the precision for a given sample size through elimination of the design effect (DEFF), which is an adjustment necessary to account for the reduction in variance introduced by clustered sampling approaches (as neighbouring households within a cluster are more likely to have similar water and sanitation service levels than households from different clusters).

Overall, these changes allowed us to generate disaggregated results for a higher number of low-income areas, while allocating a sample size to each area that was sufficient to achieve a precision of $\pm 10\%$ without requiring bootstrap resampling techniques (as in Valid International's original methodology). This has the advantage of allowing us to conduct most analyses ourselves.

In Section 2, we focus on the methodology used in the last of our 7 surveys, in Lusaka: this is essentially the methodology that we will adopt in future. In Section 7, we provide a more detailed technical description of the systematic spatial sampling methodology we have used, and discuss the pros and cons of this methodology by comparison with other candidate methodologies. In Table 1, we briefly outline the methodologies used in each of the 5 surveys, including the substantially different approach used in Dhaka and Maputo.

2. Sampling design

In what follows we describe the process we used to design our Lusaka survey, but we phrase this in question-and-response format aimed at other individuals/organisations planning similar surveys.

The approach we describe here is for large-scale household survey to assess WASH service levels in a given city or town. The approach is appropriate for urban locations in sub-Saharan Africa and south Asia, where many people live in low-income communities, where WASH service levels are typically highly deficient, and where detailed census data (i.e. a list of households with defined addresses and known number of occupants in each household) will rarely be available. The approach uses a closed questionnaire appropriate for large-scale survey: though certainly it would be possible to run more qualitative data collection (non-closed questionnaires, focus groups, etc.) in parallel. We also stress that this is a methodology for household survey, not for infrastructure survey: though again, it would certainly be possible to run an infrastructure survey in parallel.

In what follows, we outline our survey approach in non-technical terms, without getting bogged down in statistical detail (e.g. the assumptions under which the Cochran formula is valid). We are not statisticians ourselves, though this publication has been reviewed by survey design specialists with more in-depth understanding of the statistical aspects.

2.1 What is your overarching aim?

It is of course critical to have a clear understanding of your aims, and particularly of the degree and nature of disaggregation you're looking for:

- Do you want a survey of the city as a whole? Or of low-income areas across the city?

- Within the overall survey, what spatial disaggregation to you require? Administrative subdivisions of the city? Other specifically identified areas (e.g. areas in which your organisation has intervened or plans to intervene; or areas served/unerved by a particular service provider)?

- Do you require any sort of non-spatial disaggregation (e.g. by wealth level)?

In our experience, this is not a simple linear process, but an iterative process (you define your aims; you work out how you might achieve this with available resource; you then realise that your original aims were over-ambitious, and somewhat reduce those aims).

In our Lusaka survey, we decided that we wanted a survey of low-income areas⁶ across the city as a whole, with reduced sampling of just a few comparison non-poor areas, on the view that WASH service levels can be considered to be good in these areas. Specifically, we decided that we wanted sufficient power to generate separate estimates of WASH service levels **a)** for low-income area across the city as a whole (i.e. treated as a single unit) with $\pm 5\%$ precision, **b)** for disaggregated geographical groupings of low-income areas with $\pm 10\%$ precision, and **c)** for a small number of non-poor areas, for comparison purposes (more details in Section 2.6 below).

2.2 What is the available budget?

Evidently you need to have clarity about budget available for each survey. We had a maximum available budget of around \$25,000 for each survey. WSUP has historically tended to subcontract a local consultant to deliver surveys, and this budget proved sufficient for that. For our Lusaka survey, however, we had only \$10,000, and decided to manage the survey ourselves (i.e. the survey was managed directly by a member of the WSUP Zambia team, with technical support from the UK, and direct contracting of enumerators by WSUP Zambia).

⁶ In fact in Lusaka terminology, the poorest areas are called "peri-urban areas", while the areas locally called "low-income areas" are essentially middle-income by relative standards. In the text we use "low-income areas" as an international generic term: but all areas included as low-income areas in the Lusaka survey were in fact "peri-urban areas" in Lusaka terminology.

Own-management has advantages including tighter control over timelines and quality, and in future we will likely continue to use own-management where possible. Other organisations may have sufficient own-staffing for survey management, or indeed sufficient own-staffing to provide enumerators. However, we would stress that surveys need to be efficiently managed, and enumerators need to be well-trained and focused on the survey: it's not something that can be done well as a side-activity in the course of other duties.

We note that our survey budgets indicated above refer to out-sourced costs and expenses, and do not cover WSUP staff time in-country or in London. Most notably, design and management of these surveys from London has required a very significant investment of skilled staff time: probably around one person-year of work. Certainly, we would expect the staff time requirement to be substantially lower for future surveys.

Of course, it is also possible that you might decide what type of survey you want, and then estimate and seek the required budget. That would be perfectly valid, but it's not the approach we used.

2.3 What sample size can you afford?

Affordable sample size is in theory a simple calculation: divide budget by cost-per-household-surveyed. In practice, it's not so simple, because you probably don't have perfect knowledge of cost-per-household-surveyed, and you may need to make an estimate; or you may choose to put out to competitive tender with total budget specified, and see what sample sizes bidders offer. Furthermore, the relationship between sample size and survey costs is not linear: there should be economies of scale in most circumstances.

For our Lusaka survey (managed by WSUP in-house, with direct contracting of enumerators), we had a total budget of \$10,000, and estimated that cost-per-household-surveyed would be about \$8.7, giving a total available sample size of 10,000/8.7 ≈ 1155 households.

⁷ This non-specialist explanation of the meaning of "confidence interval" is not strictly correct. Strictly speaking: if one calculates the 95% confidence intervals repeatedly for different samples from the same population, then on average 95% of these confidence intervals will contain (i.e. include) the true value in that population. Hence the 95% probability is a property of a long sequence of confidence intervals computed from the same population, rather than a property of any single confidence interval. The single "95% confidence interval" emerging from a given survey has in fact a somewhat smaller probability of containing the true value.

2.4 Is that sample size sufficient to achieve the confidence and precision you need?

For our surveys, we required 95% confidence and at least ±5% precision: in other words, for any given outcome variable (say a proportion, or an average), you can be 95% certain that the real value lies within ±5% of the estimated value.⁷ [The precision is sometimes called the "margin of error".]

A confidence level of 95% is a widespread convention in survey research, so we adopt this as a routine requirement. For a given sample size and 95% confidence, the precision can be calculated using a rearrangement of the standard Cochran formula:

$$c = \sqrt{([z^2 * p(1 - p)]/n)}$$

where c = precision, z = z-value for 95% confidence interval (= 1.96), p = expected proportion of population showing the attribute (e.g. sanitation is "basic" or better), and n = sample size. For p, the most conservative (i.e. most rigorous) assumption is 0.5.

For our example:

$$c = \sqrt{([1.96^2 * 0.5(1 - 0.5)]/1155)} = 0.029 \approx \pm 3\%$$

Thus our Lusaka sample size of 1155 offers ±3% precision across the whole survey, better than our minimum requirement of ±5%.

If we rearrange the Cochran formula to calculate required sample size for 95% confidence and ±5% precision, we obtain that a much lower required sample size of 384: but following the calculations below, this would only allow a maximum of 4 separate Enumeration Areas within the sample.

2.5 How many Enumeration Areas can you achieve?

In order to calculate how many separate Enumeration Areas we could achieve within a survey with 1155 households, we accept a lower precision within each Enumeration Area (±10%),

Table 2. Enumeration Areas for the Lusaka citywide survey.

EA	Areas	Constituency	Service provider	LWSC Peri-Urban zone	Description	Sample size
1	Misisi, Kuku, Chawama, Kuomboka, Chipweny Pwenu, John Howard	Chawama	LWSC	South	Peri-Urban Area (PUA)	105
	Jack Compound	Kabwata	LWSC			
	Freedom	Chawama	Freedom Water Trust			
2	Chibolya, John Laing, New Kanyama, Linda	Kanyama	LWSC	South	PUA	105
3	Old Kanyama	Kanyama	Kanyama WT	South	PUA	105
4	George, Lilanda, Chunga, SOS Village	Matero	LWSC	West	PUA	105
	Mulobela	Matero	-			
5	Garden, Ng'ombe	Mandevu	LWSC and WT	East	PUA	105
6	Chaisa, Marrapodi, Kabanana	Mandevu	LWSC and Chaisa (LWSC and WT)	West	PUA	105
	Mazyopa	Mandevu	-			
7	Chazanga, Chipata	Mandevu	Chazanga WT and Chipata WT	West	PUA	105
8	Chainda, Kamanga, Kalingalinga, Mtendere	Munali	LWSC	East	PUA	105
9	Kalikiliki, Mtendere East	Munali	Kalikiliki WT and Mtendere WT	East	PUA	105
10	Kaunda Square Stage 1, Kaunda Square Stage 2	Munali	-	East	Low-Income Area	105
11	Kabwata	Kabwata	-	-	Middle-Income Area	105
						1,155

and use a rearrangement of the Cochran formula to work back from this:

$$n = [z^2 * p(1 - p)]/c^2$$

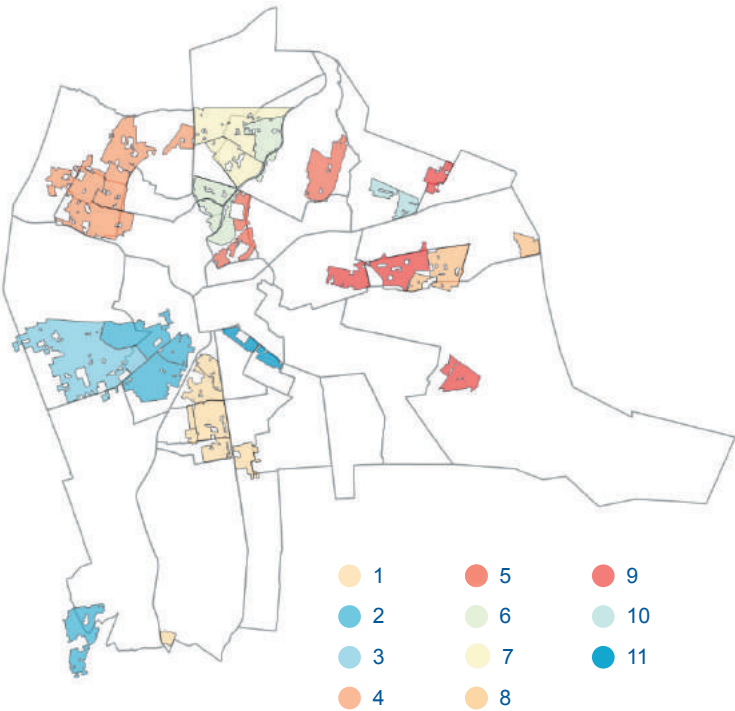
$$n = [1.96^2 * 0.5(1 - 0.5)]/0.1^2 = 96$$

Thus any Enumeration Area for which we require a separate estimate needs a sample size of at least 96. So this means we could achieve up to 12 (1155/96 = 12.03) separate Enumeration Areas within this survey. But in fact we decided that we preferred to allow some "slack" to compensate for non-response, and we eventually opted for 11 Enumeration Areas each with sample size of 105.

2.6 Identifying Enumeration Areas

Now we can relate the calculations back to reality! In Lusaka, we were able to divide the total survey area into 11 separate Enumeration Areas, allowing us to obtain estimates with 95% confidence and ±10% precision in each of the

Figure 1. Map showing the location of Enumeration Areas in the Lusaka citywide survey. Refer to Table 2 for neighbourhoods within each EA.



areas. Of these, 8 were groupings of low-income areas together covering all low-income areas in the city, and 2 were non-poor areas included for comparison purposes (see Figure 1 and Table 2). Certainly, if we had had more budget we would have ideally additionally covered all non-poor areas of the city. Note that essentially non-populated areas (e.g. airports) were excluded.⁸ These are likely to be context-dependent decisions: in some cities, low-income and higher-income populations are spatially mixed, so it is not possible to meaningfully divide the city into poor and non-poor areas.

You will need to have some reasonably accurate estimate of the population size of each Enumeration Area, and ideally a map of population density *within* Enumeration Areas. As discussed below, our method applies a population weighting to adjust for differences in Enumeration Size. But this may be unreliable if differences in population size between Enumeration Areas are very large: ideally, each Enumeration Area should have broadly similar total population size. In fact, we did not prioritise this rule in our own survey designs:⁹ but this may have been a misjudgement, and we will give greater priority to this in future (by ensuring Enumeration Areas are of broadly similar size and/or by increasing sample size in larger Enumeration Areas: see Section 8 for further discussion).

How you decide to make use of your Enumeration Areas is about practical value, not statistics. What will be most useful to local stakeholders (e.g. the utility, the municipal authority)? What will be most useful to your organisation (for monitoring interventions now starting, or for planning future interventions)?

Note that you may also consider groupings of your Enumeration Areas, as in the following (very tidy!) hypothetical example: imagine you can achieve 12 Enumeration Areas with sample size 100; imagine your city has 3 administrative divisions, each divided into 4 sub-divisions; so if you treat each sub-division as an Enumeration Area with $n = 100$, you will also obtain valid data ($n = 400$) for each of the 3 higher-level divisions.

The driving consideration in all this is the required precision. With our sampling approach, we need a sample of about 100 for each Enumeration Area, in order to generate results with $\pm 10\%$ precision. While tempting to subdivide the Enumeration Area when presenting your data, this isn't statistically acceptable: you have valid estimates for each Enumeration Area, but *not* for lower-level units within each Enumeration Area. In Lusaka, we used various criteria to define our Enumeration Areas, aiming to create areas of broadly similar population size that allowed disaggregated estimates for different administrative divisions ("Constituencies") and different zones as defined by the water utility: see Table 2. In other contexts, decisions might also be based on other pragmatic considerations, depending on the particular aims of the survey (e.g. community health worker surveillance catchments).

Certainly at this stage you might choose to define fewer Enumeration Areas, and reduce total sample size accordingly; or you might decide that you need more Enumeration Areas, and thus a bigger total sample. As noted at the outset, we describe this as a linear process, but in reality you're likely to circle back at different points in your design.

2.7 Dividing Enumeration Areas into PSUs

Having defined our Enumeration Areas, we then used the Centric Systematic Area Sampling approach as developed by Valid International (of which more in Section 7). To do this, you need to delimit each Enumeration Area on a Google satellite map (or other high-resolution map): you should delimit each area in ESRI shapefile format, probably using commercial software like ArcGIS or free software like QGIS. For each Enumeration Area, you then need to grid the area with a total number of grid-squares corresponding to sample size for each Enumeration Area (105 in our Lusaka example). In statisticians' terminology, each grid-square is a Primary Sampling Unit (PSU). To grid each area, we used a web app developed for us by

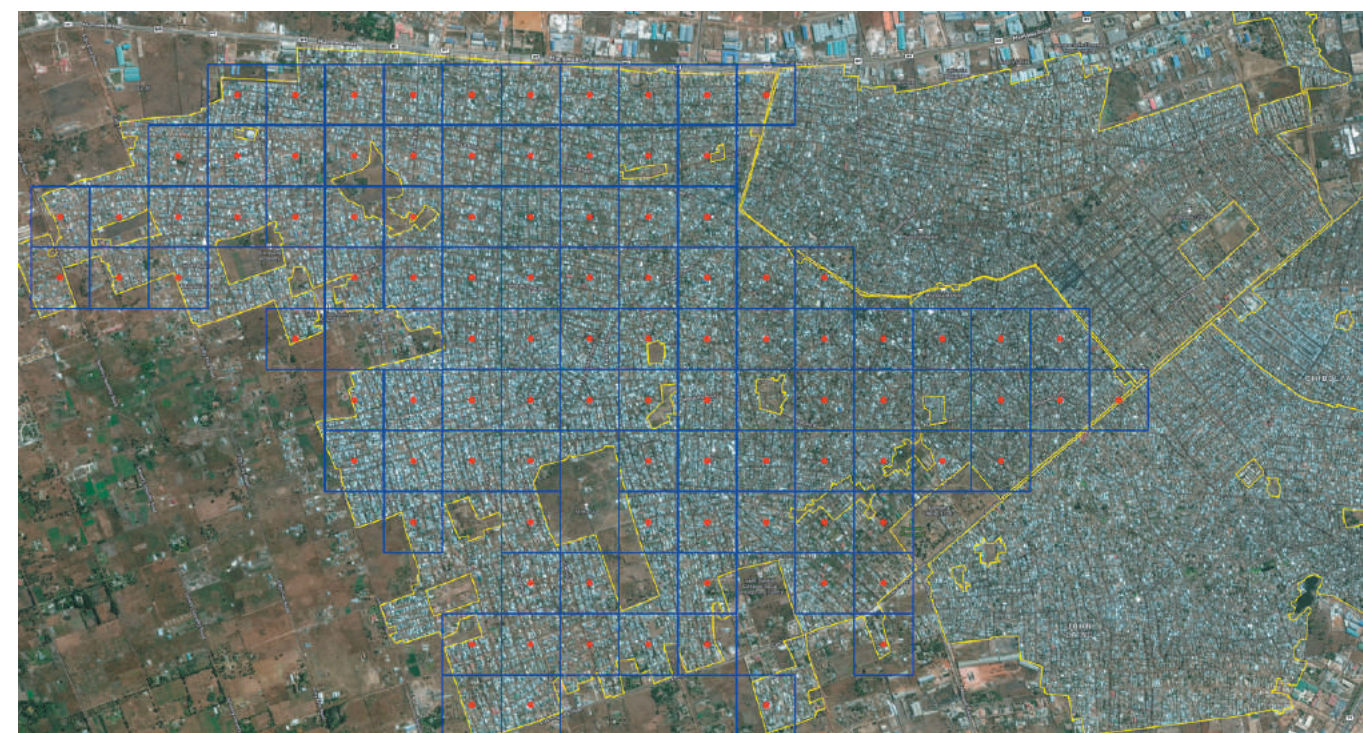
Valid International, which is currently freely available at <http://apps.validmeasures.org/wsups/> [> select tools > spatial sampling]. An example grid for one of the Lusaka Enumeration Areas is shown in Figure 2. The Centric Systematic Area Sampling approach involves selecting the single household closest to the centre of each grid-square (PSU), as further discussed in the next section.

2.8 Household selection protocol

The single household at the centre of each grid-square is the Secondary Sampling Unit. Specifically, our sampling protocol required the enumerator to find the closest residential building with an adult female willing to take part in the survey. Female respondents were required because the questionnaire includes gender-specific questions (e.g. around women's perceptions of safety when using toilets, around menstrual hygiene management). Ideally the respondent was the

female head of household, but if there was no female head of household or she was not available, another adult female member of the household was asked to participate. If no adult woman was in the house at the time of the survey, or if an adult woman was present but didn't wish to participate in the survey, enumerators applied a defined rule-set to replace the household with another one nearby, within a maximum distance ("tolerance") of half a grid-square width. In the rare event that no valid household was found within that distance, each situation was analysed remotely by the WSUP team using satellite imagery (Google Maps or Bing Maps), and specific guidance was provided (the maximum distance was increased, or that grid-square was excluded from the sample; the latter only in situations in which no households could be found, for example in market areas which can seem residential from satellite imagery).

Figure 2: Example Enumeration Area from Lusaka survey, showing gridding to create 105 grid-squares for sampling.



⁸ Excluding non-populated areas (airports, large roads, rail stations and tracks, green areas, water bodies, markets, industrial and institutional areas) is essential, because inclusion of such areas will pointlessly increase enumerator's time and thus survey costs, and may lead to sampling biases.

⁹ Factors of difference in estimated population size between largest and smallest Enumeration Area were Dhaka 8x, Maputo 2x, Nakuru 10x, Antananarivo 5x, Mahajanga 8x, Toliara 3x, Lusaka 9x

3. Data collection tools and processes

This section summarises data collection tools and the process of survey delivery.

3.1 Questionnaire

The questionnaire used by WSUP for these surveys has over 100 questions, mostly with closed response, around water, sanitation and hygiene including menstrual hygiene.¹⁰ A specific module is included for wealth assessment. The questionnaire varied slightly from one survey to another, in part because of city-specific requirements, in part because of our own evolving understanding of which questions worked well (see also Section 5.1).

The full questionnaire used in Lusaka can be downloaded from the WSUP website (<https://www.wsup.com/research/citysurveys/>). In summary, the questionnaire comprises the following questions:

Informed consent: This is as required by EU data protection regulations (GDPR), and simply by ethical best practice.

Identifier information: GPS coordinates, administrative area, plot number, respondent's gender and land ownership status. [For discussion of data protection implications when reporting survey data which might potentially be identifiable to a particular household, see Section 6.]

Basic sociodemographic information: Number of household members, their age and sex; number of households living in the same plot and number of people in each of these other households.

Water: Includes questions around types of source and their accessibility, acceptability, affordability, availability and quality; additional questions cover storage and treatment practices, and willingness to invest in improving water supply.

¹⁰ Note that the length of the questionnaire can influence decisions around sample size overall and per grid-square. If the questionnaire is very short, it could make sense to sample more households within each grid-square: it wouldn't make much sense for enumerators to spend half an hour finding the centre point of the square, and then only spend 10 minutes on an interview before moving to another square. In our case, however, the questionnaire was quite long: so the limiting factor was the length of the interview, not the time spent identifying the eligible household in that grid square. So it made sense to limit the number of households per grid square to one, which removes the design effect that would occur if several households were sampled in a square. See also Section 7.

Sanitation: Includes questions around the type of facility, usage arrangement, accessibility, acceptability, affordability and quality; additional questions relate to the service provider and faecal sludge management.

Hygiene: Questions covering menstrual hygiene management and handwashing practices.

Poverty assessment: This is based on the Poverty Probability Index (PPI) (<https://www.povertyindex.org/>), a wealth assessment tool based on household characteristics and asset ownership, with country-specific adjustments and calibrations publicly available. This gives a probability that the household interviewed falls below a given definition of poverty.

Observational section: Enumerators assess the sanitation, water storage and handwashing facilities and take pictures.

3.2 Additional data collection for assessment of safely managed services

In order to assess the proportion of people with safely managed water and sanitation services, as defined by the JMP, household questionnaire data was complemented with additional data (purpose-collected, or existing third-party data) on water quality and on proportion of faecal waste safely treated. [For details of data collected in each city, see Table 3.] For Maputo and Nakuru, budget for additional water quality testing was available.

3.3 Survey delivery

All surveys were conducted using an online data collection platform, mWater (<https://www.mwater.co/>) which we have found to be highly effective and user-friendly. Data entry is directly to an Android device (phone or tablet), via a questionnaire created in mWater and downloadable to the mWater Surveyor app on each phone.

In the Lusaka example, we contracted 20 enumerators, who were trained and deployed in May/June 2018, right after the rainy season. Enumerators were provided with a list of GPS coordinates, linked to Google Maps, for each target household location at the centre of each grid-square; they were also given written guidance notes and maps of their Enumeration Area/s with each grid-square indicated, including specification of maximum distance of the selected household from centre of grid-square. Enumerators were expected to complete at least 6 surveys per day. Most of them used their own Android phones with the mWater Surveyor app and Measure Distance Map app (initially we supplied low-cost Android tablets, but these showed issues with battery life and GPS accuracy). Measure Distance Map allows you to

check how far you currently are from a specific GPS location, and was used by enumerators to check that the household they were surveying was within the acceptable tolerance from the target GPS location.

In Lusaka, as in other locations, we experienced two practical problems with survey delivery. First, as noted, there were problems associated with short battery life and poor GPS accuracy of some Android devices. Second, enumerators had difficulties accessing non-poor households, often in gated condominiums. These issues are discussed in more detail in Section 8.

Table 3. Data collected for estimation of safely managed services. For water, distinguishing between the service levels “basic” and “safely managed” requires water quality data, which can be tested at household level as part of the survey, or derived from existing city-level data. For sanitation, distinguishing between “basic” and “safely managed” requires data on what proportion of excreta is safely treated and disposed, whether on-site or after transport in a truck or along a sewer; this necessarily requires city-level data. This table summarises data used for distinguishing between basic and safely managed. [This table is just to give a brief indication of the types of data used: we don't provide full references or methodology details, but are happy to provide these on request.]

Service	Dhaka	Maputo	Nakuru	Madagascar	Lusaka
Water	– No reliable data available, except water quality data for one specific slum	– Water quality testing in 64 HHs – Interviews with regulator, utility and municipality	– Water quality testing in 100 HHs – Interviews with utility	– Existing WSUP data on quality of water from kiosks – Existing WSUP data on water quality in Mahajanga and Toliara	– Data from utility's monitoring system – Water quality data from CDC study exists however this isn't yet publicly available. – Ministry of Health data on borehole water quality
Sanitation	– 2016 Shit Flow Diagram (World Bank)	– 2014 consultant assessment of FSM market – 2015 review for creation of Shit Flow Diagram – 2014 World Bank (WSP) study of sanitation in Maputo – Interviews with municipality	– Focus groups with community members – Focus groups with manual pit emptiers – 2015 Shit Flow Diagram	– 2018 review of sanitation in Antananarivo – 2017 report on new treatment facility in Mahajanga	– Waste flows data generated by Lusaka Sanitation Plan

4. Quality assurance and data cleaning

This section outlines the process used for quality assurance and data cleaning.

4.1 Quality assurance

A local supervisor oversaw enumerators in the field. Any queries and problems were reported to the supervisor at the end of each day and that person contacted the survey management team in London for any clarification on a daily basis. If there was any indication of issues, feedback was given to the supervisor. The main issues encountered were households surveyed outside of allowed tolerance or outside of the intended grid-square; no household found within a grid-square, or occupants unwilling to take part in survey or allow photographs of facilities; and issues with question wording and design that emerged as the survey proceeded.

Tolerance (maximum acceptable distance of the selected household from the centre of the grid-square) was assessed by measuring the distance between the grid-square-centre GPS coordinates and those of the sampled household, in an Excel spreadsheet. The accuracy of locations was double-checked from London daily (mWater records the GPS location of the survey, and this can be checked for each enumerator or all enumerators). Households beyond the tolerance limit were rejected and the enumerator was required to find another household within the tolerance; in a few cases, WSUP gave approval to skip a grid-square, because no eligible household could be found within the tolerance.

Each data entry (i.e. each set of responses from a single household respondent) was only marked as “Final” on mWater once that entry had been approved by WSUP.



Image: Kanyama community market, Lusaka. Credit: Gareth Bentley

4.2 Data cleaning

The data cleaning protocol was as follows:

1. Data collected through mWater and marked as “Final” by the supervisor is downloaded in CSV format.
2. The GPS coordinates are plotted on a map using Google Earth.
3. In (rare) cases where enumerators sampled more than one household within the same grid-square, the closest household to the grid-square centre was included in the survey.
4. Answers in the “other (please specify)” category were classified into the main categories when possible.
5. Typos were corrected (wrong PSU numbers, wrong Enumeration Area, mistakes in recording units, etc) based on the location of the PSUs.
6. The data-cleaning process was documented, and each action specifically recorded.
7. The final database was saved in CSV format for analysis.

5. Data analysis

The analysis of the 7 surveys was done jointly: i.e. all data is transcribed into a single CSV file. This transcribing process was conducted in Excel using a formula-driven sheet based on lookup tables that dynamically draws the appropriate questions (taking into account questionnaire variations) into one dataset with consistent headers. Subsequent data analysis was performed using a combination of R, Stata and Excel.

5.1 Indicators and indicator mapping

The questionnaire was created after identification of the derived indicators that we required. These indicators have been developed specifically for the WSUP citywide surveys, but most are closely aligned to the existing JMP standards, the SDGs and the Human Right to Water and Sanitation.

The questionnaire used has evolved between the first survey in Dhaka and the last in Lusaka, in part because of city-specific requirements and adaptations, in part because of our own evolving understanding of which questions worked well. Most of the indicators from the final indicator set could be obtained for Lusaka, but there are some data gaps in the other surveys. Furthermore, where the precise question asked differed among surveys, the corresponding responses need to be aligned for consolidated analysis: this could just be simple minor differences in question wording (e.g. “How long does it take to go there, get water, and come back?” versus “How long does it take to go to the source, queue, get water and come back to the dwelling?”) or it could be a more complex issue.

For example: we ask all respondents how many hours-per-day and days-per-week water is available from their primary water source. Following the initial surveys, this question was branched into two questions, covering the rainy season and post-rainy season, to gather better data on seasonal variability. In order to produce a consistent dataset, these two values were then weighted to provide overall annual figures which

are consistent with the previous question asked. This whole process is formula-driven and based on logic tables in Excel, so all question-mapping steps are easy to trace and easy to modify.

As part of the data analysis, questionnaire responses need to be mapped to corresponding indicators. Key to these surveys was assessment of household position on the JMP water and sanitation ladders. For the water ladder (for example), the household’s position was determined based on the main source of drinking water (and its classification as either an improved or unimproved facility); how long it takes to collect water; hours of availability; and location (either on-premises or off-premises). For distinction between “basic” and “safely managed”, additional data is required (see Section 3.2): so the household rating is then combined with water quality data at the Enumeration Area level, to generate the final assessments of proportion in each ladder category within each Enumeration Area. A full list of the indicators derived from WSUP citywide surveys is available for download from the WSUP website (<https://www.wsup.com/research/citysurveys/>).

5.2 Weighting by population size

As already noted, the sampling strategy was not proportional to population size: in other words, a single sample was obtained from each of the 105 grid-squares, without regard to population density in each grid-square; and at the next level up, the different Enumeration Areas comprising the total sample differed in total population size. This means that **a)** in order to generate aggregate data for each Enumeration Area, the individual responses for each grid-square need to be weighted by the estimated population of that grid-square, and **b)** in order to generate aggregate data for the whole city, the estimates for each Enumeration Area need to be weighted by the population of that Enumeration Area. For example, imagine we have a city with three Enumeration Areas A–C with population sizes 200,000 in A, 400,000 in B and 800,000 in C; and imagine that *proportion of population with basic sanitation* (as estimated from our surveys)

Table 4. Approach used in the different surveys to obtain population size estimates for each Enumeration Area and grid-square.

Dhaka Population data in slums obtained from 2014 “Census of slum areas and floating population” carried out Bangladesh Bureau of Statistics. Wider population data was obtained from WorldPop (www.worldpop.org.uk).
Maputo 2007 population data were obtained from the National Institute of Statistics, down to bairro level (bairros are urban sub-districts with population in the range 3,000–50,000). [Preliminary results from the 2017 census were released in August 2018, and when the full data becomes available we may be able to more accurately weight our analysis.]
Nakuru To estimate population sizes in each Enumeration Area, 2013 population data were obtained from the utility NAWASSCO, down to “low-income community” level; current population sizes were estimated assuming a constant growth of 1.045% per year since 2013. A 2009 census was used to estimate population in areas not covered by the 2013 census.
Madagascar 2009 data from INSTAT/DDSS based on consultations with local community leaders (Chefs Fokontany).
Lusaka 2010 data held by the utility.

is 60% in A, 80% in B and 70% in C; weighting by population size, we can estimate *proportion of population with basic sanitation* across the whole city as:

$$(200,000/1,400,000)*0.6 + (400,000/1,400,000)*0.8 + (800,000/1,400,000)*0.7 = 0.71$$

The challenge is of course to obtain reliable data on population size in each grid-square and each Enumeration Area. The types of data source used varied among surveys: see Table 4 for detail. In almost all cases obtaining reliable population data was challenging. Obtaining the best available data typically required extensive searching of online sources and/or efforts to obtain data held by local institutions; often the best available data was outdated and approximate, and we certainly recognise this uncertainty as a significant limitation of this survey methodology (see Section 8)

In Lusaka (and all other surveys except Dhaka), population within each grid-square was typically estimated using existing population estimates for sub-areas of the Enumeration Area. So imagine 105 grid-squares within a given Enumeration Area; imagine that the Enumeration Area comprises 3 sub-areas A-B-C with known populations of 10,000, 20,000 and 30,000 respectively (so total population 60,000); imagine that A is covered by 20 of the 105 grid-squares; so within A, average population per grid-square is 10,000/20 = 500; so to obtain the aggregate

value of variable x for this Enumeration Area, each value from a grid-square in A is weighted by a factor of 500/60,000 (and similarly for the grid-squares in sub-areas B and C).

5.3 Data analysis summary

In summary, the key steps in data analysis are as follows:

- Input the raw data from each survey into a single spreadsheet (one tab per city survey, where each row is a single respondent’s set of responses).
- Automated processing within Excel, based on logic tables, driving creation of a single tab containing a consolidated and cleaned dataset which is then exported as a CSV file for further analysis. This generates indicators which can be assessed at the household level.
- A separate CSV file is compiled which holds data at the city level (such as estimates of the proportion of sludge from on-site facilities which is safely treated), required for determining the proportion of people in the city with “safely managed” sanitation.
- A weighted analysis is conducted (in R, Stata, and also in Excel as a final check) to apply the population weightings required in order to generate indicator values for each Enumeration Area and at city level.

6. Survey reporting and visualisation

In coming months, we will make full data from these surveys publicly and freely available. At time of writing, we are still assessing exactly how we will do this.

We may release map-based data visualisations via a visualisation platform like Qlik Sense or Tableau, or we may publish data in simple Excel format. We will not publish full raw data, primarily because of issues around data confidentiality for

respondents. Notably, any presentation of data (whether in tabular or mapped form) needs to ensure that individual households are not identifiable in any way. In the case of mapping, this may require “de-focusing” (e.g. showing data only for clusters of 5 respondent households, these not individually plotted on the map).



Image: Primary schools, Ambohimandroso, Madagascar. Credit: Johan Mottelson

7. Systematic Spatial Sampling versus other sampling approaches

Throughout the text, we have described our approach without in-depth consideration of its statistical validity. Here, we provide a more detailed technical description of the systematic spatial sampling methodology we have used, and discuss its pros and cons by comparison with other candidate methodologies.

The approach used in Lusaka can be more precisely described as a *map-based systematic spatial sampling approach without clustering, and with posterior weighting for population size*.

This approach, like the other approaches outlined below, is *area-based* as opposed to *list-based*. A list-based approach requires a reasonably accurate census (a “sampling frame”): i.e. a list of all households in the city, with known number of people in each household, and with readily locatable addresses. If a census is available, true random sampling can be used (random selection of *n* households from the list), and this immediately eliminates the two big challenges of area-based sampling, i.e. dealing with variations in population density, and achieving true random selection. However, up-to-date censuses were not available in any of the cities in which we carried out surveys and are very rarely available in developing contexts. [For small survey locations, e.g. a particular slum, an option is to actually carry out a census, and then use that as sampling frame: but clearly that is not likely to be a viable option for citywide surveys.]

For area-based sampling, the following are the main options available within any given Enumeration Area:

1) Map-based random spatial sampling, no clustering: For a sample size of 100, the survey designers identify 100 randomly selected GPS coordinates within the Enumeration Area, and include within the sample the eligible household closest to each point. In this approach, there is a likely bias due to higher probability of inclusion of households in areas of lower population density, and correction of this bias is challenging: even if

variations in population density can somehow be estimated throughout the area, weighting would then require sophisticated and complex spatial statistics methods. We therefore consider that this method is problematic except in locations with highly uniform population density, if such locations exist. A further issue is that, if we identify points completely at random, then in some Enumeration Areas we may be “unlucky” and miss out important parts of the area, generating unrepresentative results.

2) Map-based random spatial sampling with clustering: For a sample size of 100, the survey designers identify say 25 randomly selected GPS coordinates within the Enumeration Area, and then use some sort of quasi-random “random walk” protocol to select 4 households around each point (e.g. “*spin a bottle, walk in the direction of the bottle and select the first household on the left, then keep walking in the same direction and identify the next household 3 doors down, then again, then again*”). Again, with this approach there is a likely bias due to higher probability of inclusion of households in areas of lower population density, and in addition quasi-random processes of this type are susceptible to multiple biases. Clustering has the potential advantage of reducing enumerator travel time and therefore reducing survey cost. But in the WASH context it is problematic because of strong suspicion that WASH service levels will tend to show strong spatial correlation: this is particularly the case for services that require external/public investment in community infrastructures, such as water networks.

3) Map-based systematic spatial sampling, no clustering [as used by us in Lusaka, and as described in this Research Brief]: For a sample size of 100, the survey designers divide the Enumeration Area into 100 grid-squares, and include within the sample the eligible household closest to the centre of each square. This approach makes sure that all parts of an area are included. If population size within each grid-square can be estimated with reasonable accuracy, weighting by population size can then be applied to create aggregated indicator

estimates for the Enumeration Area as a whole: this can be done using straightforward conventional (non-spatial) statistical analysis.

4) Map-based systematic spatial sampling with clustering [as used by us in early surveys]: For a sample size of 100, the survey designers divide the Enumeration Area into say 25 grid-squares, and then use some sort of quasi-random process to select 4 households in each square. Again, if population size within each square can be estimated with reasonable accuracy, weighting by population size can then be applied to create aggregated indicator estimates. This approach can potentially reduce survey cost, but we see no strong advantage in the citywide survey context, and it has probable disadvantages related to the likely spatial correlation of WASH service levels. However, it may make sense in situations in which the survey uses a short questionnaire that makes it easy to cover many households in a short time. For example, if we wish to estimate the effect of an intervention on diarrhoea, then it could make sense to go to the centre point in a grid and systematically enrol a fixed number of households from there. This approach would make more sense for variables (e.g. hygiene behaviours, or diarrhoea morbidity) that we would expect to be less spatially clustered than WASH service levels.

An additional alternative would be a quasi-random “random walk” method with starting points visually identified by the enumerator on the ground (e.g. an intersection, a shop or a church), subsequently applying some sort of “spin a bottle” protocol. The problem with this is that often supervisors have too much freedom to choose starting points, and enumerators too much freedom to choose which way to walk and which house to enrol. It is very difficult to identify a good starting point: ideally the starting point should be random, but preference is often given to starting points more centrally located in an area. Our method should be better, as **a)** the choice of the starting point is effectively random in space (since the grids are laid arbitrarily), and **b)** the rule for enumerators to



Image: Tertiary network construction in Charmanculo, Mozambique.

enrol a house are strict (“the house nearest to the centre point of the grid-square”).

Finally, a possible alternative approach would be to select a subset of areas within the city, and exhaustively sample all households within each area; however, it seems unlikely that this approach would be effective for citywide sampling of WASH service levels.

8. Method limitations

This section identifies the major limitations of the methodology we used, and discusses ways in which we may modify our approach in future. Probably the major limitations of our approach relate to population density weighting, as detailed in the first two paragraphs below. As noted in the previous section, weighting for population density is necessary in any area-based (as opposed to list-based) approach: but most surveys of this type will be in contexts in which there is no good recent census providing a basis for list-based sampling.

We faced difficulties in obtaining accurate up-to-date population density data with sufficiently fine spatial resolution. This needs to be accepted as an unavoidable difficulty of area-based sampling methods for citywide surveys: survey designers need to access the best data available from existing censuses, population density maps and other sources. An alternative would be to use some sort of roof-counting algorithm applied to satellite imagery, to obtain own-estimates of population density; but there are multiple challenges to doing this effectively at the city scale, particularly in informal settlements.

We made a questionable design decision in allowing Enumeration Areas to have very different total populations.¹¹ In some surveys the population size of the largest Enumeration Area was 10 times the population size of the smallest Enumeration Areas.¹² This means that, if the sampling within a large area by chance gives an inaccurate result (95% confidence is not 100% confidence), then that inaccuracy will be heavily transferred to the aggregate estimates for the city, because of the high weighting of large areas. We are currently carrying out analyses of our

data to assess whether this issue was likely to have caused significant inaccuracy. This issue could be handled by either **a)** ensuring that all Enumeration Areas have broadly similar population sizes and/or **b)** by increasing sample size in larger areas to increase confidence in estimates obtained for those areas. Evidently, increasing sample size in larger areas means either increased financial cost, or reduced precision in smaller areas, or a reduction in the number of Enumeration Areas the survey can cover.

Other significant difficulties faced in survey design and data collection included the following:

Identifying low-income areas was challenging in some cities. In Dhaka, many low-income areas are very small and scattered throughout middle/high-income areas, in some cases just a small cluster of low-income households sandwiched between high-rise developments; furthermore, the available maps were severely out-of-date, so we had to rely on a more recent census listing of slum areas from the Bangladesh Bureau of Statistics; in some cases we had challenges locating listed slum areas, because they were very small and/or listed in the census with names not known to people living in that area. In Antananarivo, the challenges were of a different type. Most of the city's population can be considered low-income, with low-, middle- and high-income households often co-existing in a close patchwork: in other words, it would often not be meaningful to identify "low-income areas". So if the aim is to focus on low-income populations, survey designers need to consider carefully how this can be best achieved in any given city [We also note that, in some situations, homeless populations would need to be considered].

¹¹ Our primary drivers were to identify Enumeration Areas that a) provided a pragmatically useful disaggregation of the city to WSUP and other actors, and b) ensured sufficient sample size in each area to give 95% confidence and $\pm 10\%$ precision. In retrospect, however, we should probably have given more attention to population size in each Enumeration Area, to strengthen the citywide estimate.

¹² The major concern is not with unusually small areas, but with unusually large areas. Unusually small areas do not risk "corrupting" the citywide estimation; unusually large areas do.



Image: Peri-urban area, Madagascar. Credit: Johan Mottelson

Accessing higher-income communities was often difficult, because many higher-income households are in gated compounds with controlled access. If it is reasonable to assume that the great majority of higher-income households have high water and sanitation service levels, this is not a major concern in practical terms, because if that assumption is correct it's pointless to waste sample size on non-poor households. But in most cities this assumption is probably not watertight: early analysis of our data suggests that in some of the higher-income communities included in our surveys, a significant minority of households did not have high-quality water and sanitation services.

Related to the previous point, we are not entirely confident that the PPI wealth assessment module included in our questionnaire gave reliable results. The PPI modules are nominally designed for both rural and urban communities, and are custom-designed and calibrated for each country; this is probably the best standard system

available, and we will likely use it again. But we have some doubts about its accuracy in distinguishing between wealth quintiles in urban contexts.

Finally, we experienced issues with the GPS accuracy and battery life of Android devices (tablets and phones). This was an irritating and time-consuming problem, but probably does not have significant implications for our findings. GPS inaccuracy, and delays in making satellite links, was a common problem, particularly with cheaper and older devices. We note that the mWater website has useful guidance on device selection, set-up and use. We also note that the mWater Surveyor app very usefully shows the user the current accuracy of GPS location.

9. Concluding remarks

We have described a procedure for design and delivery of surveys to assess WASH service levels across cities in developing contexts, with particular focus on low-income communities.

We currently consider that the approach used (map-based systematic spatial sampling with posterior weighting for population size) is essentially the right approach for our purposes, and we expect to use this approach in future surveys. However, we highlight the challenges of obtaining high-quality population density estimates, as required for this approach. We also draw attention to a possible disadvantage of our design (wide variation in the population size of different Enumeration Areas), and we may approach this differently in future surveys. Finally, we note that design of surveys of this type is complex, and design decisions are of course dependent on the particular aims of any given survey, and on the constraints (e.g. budget constraints, population data availability constraints) in any given context.

Image: Assessing water infrastructure in Lusaka, Zambia. **Credit:** John Healey



Credits and acknowledgements:

This publication was written by Nerea Ajuriagogeascoa, Guy Norman, Wolf Schmidt and Jonathan Stokes. The surveys were designed and managed by NA and JS, with technical support from Valid International in design of the methodology and data analysis protocols used in the Dhaka survey, subsequently adapted for the other surveys. Technical advice on data analysis was provided by WS. This paper was reviewed by Kelly Baker.

We thank WSUP country teams, the mWater team and other partners who have contributed to the delivery of these surveys.



**Urban Sanitation
Research Initiative**