Working paper

Practical reflections on conducting survey research in cities in low and middle-income countries

Felix Agyemang, Sean Fox, Jessica Hagen-Zanker, Hussain Bux Mallah, Rashid Memon, Jeffrey Paller, Levi John Wolf

July 2023

Abstract

Researchers, planners and policymakers in low- and middle-income countries often lack up-to-date and representative data about urban populations. Censuses are infrequent and expensive, while periodic surveys generally rely on sampling frames derived from censuses. In contexts of rapid population change, this can limit the quality and value of the data collected. Moreover, even well-executed survey exercises can result in systematic biases or exclusions of certain groups.

Based on more than 50 years of combined experience of conducting survey research in a wide range of contexts, we offer practical reflections and suggestions for overcoming some of these problems. We discuss issues related to research design, sampling in cities without reliable census data or lists, practical issues associated with the design of survey instruments and the process of enumeration.

We conclude with ideas for reducing the time and costs associated with gathering high quality data in future survey exercises.

How to cite: Agyemang, F., Fox, S., Hagen-Zanker, J., Mallah, H. B., Memon, R. Paller, J., Wolf, L. J. (2023) Practical reflections on conducting survey research in cities in low and middle-income countries. ODI Working Paper. London: ODI (https://odi.org/en/publications/practical-reflections-on-conducting-survey-research-in-cities-in-low-and-middle-income-countries/)

Acknowledgements

Many thanks to Marcela Rubio (ODI) and Kristian Hoelscher (PRIO) who provided helpful comments to an earlier version of this paper.

About this publication

This paper was co-developed as part of the research project "Quantifying Cities for Sustainable Development: Transforming urban data collection to support research and policy making in developing countries", funded by UK Research and Innovation (Ref: ES/R009848/1). We also drew on lessons from the Aligning Migration Management and the Migration-Development Nexus (MIGNEX) project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 770453.

About the authors

Felix Agyemang [ORCID: 0000-0001-6000-8790] is a Lecturer in Data Science at the University of Manchester.

Sean Fox [ORCID: 0000-0002-4965-6823], is Associate Professor in Global Development at Bristol University.

Jessica Hagen-Zanker [ORCID: 0000-0003-3107-9789] is a Senior Research Fellow at ODI and a Global Fellow at Peace Research Institute Oslo (PRIO).

Hussain Bux Mallah [ORCID: 0000-0002-9415-411X] is a Research Associate at the Collective for Social Science Research in Karachi.

Rashid Memon [ORCID: 0000-0002-8543-3301] is an Assistant Professor at the Social and Economic Survey Research Institute, Qatar University.

Jeffrey Paller [ORCID: 0000-0001-6983-065X], is an Associate Professor of Politics and the University of San Francisco and Researcher at the Governance and Local Development Institute at the University of Gothenburg.

Levi John Wolf [ORCID: 0000-0003-0274-599X], is a Senior Lecturer in Quantitative Human Geography at the University of Bristol.

Contents

Acronyms

CAPI	Computer-assisted personal interviewing
DHS	Demographic and Health Surveys
EA	enumeration areas
GLD	Governance and Local Development Institute
HIES	Household Income and Expenditure Survey
ICT	Information and communication technology
IGC	International Growth Centre
LMICs	Lower-and Middle-Income countries
LSMS	Living Standards Measurement Survey
OSM	Open Street Map
SLRC	Secure Livelihoods Research Consortium
UN	United Nations
USAID	United States Agency for International Development

1 Introduction

Fast-growing cities in low and middle-income countries are critical arenas for global efforts to tackle poverty and establish environmentally sustainable economic growth. The global urban population is projected to grow by more than 2 billion by 2050, with most of this growth occurring in Africa and Asia (United Nations, 2019). Developing evidence-based policies for managing this growth is an urgent priority given the global implications of these demographic trends (Elmqvist et al., 2019; Nagendra et al., 2018). Yet even basic data about the composition and characteristics of populations are lacking in many of the world's fastest growing cities. This data deficit not only inhibits academic research but also undermines the efforts of urban planners and policymakers to pursue evidence-based action in order to tackle the climate crisis, public health emergencies, and achieve inclusive development.

There are several reasons for the absence of reliable demographic data in urban areas. Many of the world's fastest growing cities are in low-income countries with weak statistical capacity (Wardrop et al., 2018). As a result, census data - considered the gold standard of demographic data and an important baseline for other forms of data collection - may not be conducted at regular intervals or may be considered unreliable due to uneven geographic coverage, failure to capture certain subpopulations, or corruption (Carr-Hill, 2013; Kuffer et al., 2022; Wardrop et al., 2018). For example, both Pakistan's 2017 national census and Nigeria's most recent national census in 2006 proved highly controversial with accusations of corruption and intentional distortion of population counts for political purposes.

Even where reliable census data has been collected, it may not be available at spatial units that support meaningful analysis within individual cities, or it may be out of date given the pace of demographic change. For example, between the 2002 and 2012 censuses, the city of Dar es Salaam in Tanzania nearly doubled in size, from 2.3 million to 4.4 million,while the Indian city of Chennai grew from 6.6 million to 8.6 million over a similar period.

In the context of such rapid change there is a need for more frequent and reliable data collection. Sample surveys are a common means of filling data gaps between census rounds and addressing a wide range of questions of interest to scholars, policymakers and practitioners, such as issues related to public health, the economy, and politics. Well known examples include the World Bank Living Standards Measurement Study, the Demographic and Health

Surveys (DSM) funded by USAID, and citizen attitude 'barometer' surveys (e.g. Afrobarometer, Arabbarometer, Latinbarometer) covering a wide range of topics related to society and politics. These are excellent resources for researchers and policymakers, but they have some key limitations when it comes to urban research: they tend to rely on census-derived sampling frames, do not provide information that is representative at the city or sub-city scale, and are infrequent.

More recently, researchers and policymakers have increasingly adopted 'big data' approaches that rely on harvesting information from ubiquitous digital footprint data from information and communication technologies (ICTs) such as mobile phones, GPS-enabled objects, and personal computers. While a valuable addition to any analyst's toolkit, it is challenging to ensure the representativeness of this data, which is necessary for inferences about population characteristics to be unbiased (Harford, 2014; Hargittai, 2015; Hilbert, 2016; Kaplan et al., 2014; Thakuriah et al., 2017). Moreover, researchers are limited by the data made available, which may not suit the needs of researchers, analysts, or other end-users. Given the limitations of relying on data harvested from ICTs in terms of sample representativeness and scope of data collected, any attempts to meaningfully collect data about urban population in Lower-and Middle-Income countries (LMICs) in the short and medium term will rely on surveys, with a sampling strategy that is specifically adapted for urban settings. .

For those interested in generating original and representative data, we offer reflections on the practice of survey research in cities in LMICs. We build on a growing body of published research and our collective experience of conducting surveys at various geographic scales in cities across a wide range of countries and contexts. This is not a comprehensive guide, but rather a collection of key issues and insights that may help improve the efficiency and quality of future survey exercises by others, with a particular focus on developing sampling approaches for urban areas without censuses.

We focus on developing research designs that accurately reflect the urban environment, generating representative samples, and implementing effective survey instruments in urban contexts. We conclude with some innovations to enhance the efficiency and quality of future urban research, including the use of street-intercept surveys, short-format survey instruments, and statistical reconstruction techniques.

2 Research design: who, where, and at what level of analysis?

The rapid urbanisation of societies in low- and middle-income countries presents numerous challenges for researchers due to social and institutional complexity in these spaces, and requires creative strategies to collect data (Auerbach et al., 2018). Cities and their neighbourhoods tend to be ethnically and religiously diverse, politically pluralistic, administratively multi-sectoral, and economically varied. Additionally, neighbourhood boundaries are often fluid and contested.

Researchers must decide their unit of analysis (e.g. individual, street, neighbourhood, city, metropolitan region) in the context of nested and embedded decision-making and service provision procedures, requiring necessary trade-offs. For example, capturing a representative sample of the population of a specific neighbourhood may not be suitable if the intended application is to inform metropolitan-wide plans or policies. Moreover, surveys play an important role in capturing the lived reality of urban life, but the complexities often also benefit from qualitative methodologies to ground-truth survey instruments and interpret the results.

Research designs often require revision in light of resource constraints and contextual realities. It is therefore useful early in the process to take stock of available or required *resources* (including time, money, expertise, staff availability and capacity) and *context* (including linguistic barriers, security threats, or the sensitivity of the research), which are often amplified in cities. Failure to consider resource requirements and context in detail can result in a mismatch between resources awarded and those required to execute a survey exercise as imagined.

With this in mind, three questions should guide the research design: (1) *who* needs to be surveyed?; (2) *where* can the population of interest be found?; (3) at *what* level of analysis is the data being gathered? In practice, research design is an iterative process, with research objectives and questions often evolving in both the development and implementation phases (Kapiszewski, MacLean, and Read, 2022).

2.1 Who needs to be surveyed?

The who of the research - the population of interest - needs to be defined as tightly as possible with explicit consideration of where this population will be found (Table 1). Defining the population of interest requires identifying a *specific* location of interest (e.g. a neighbourhood, district, city, etc.). To help find this population, we can also consider the *type* of place where they live or work (e.g. informal settlements, gated communities, peri-urban areas, etc.). Identification of these defining characteristics of a population of interest informs the sampling strategy and provides the foundation for making claims about the representativeness of the resulting sample.

Table 1 Defining the population of interest

	People	**Type of place**
Type	e.g. youth, women, ethnic group, migrants, etc.	e.g. informal settlement, gated community, peri-urban area, etc.
Location	Street, neighbourhood, city, region	
Administrative Unit	Zone, District, City, Region, Province, Metropolitan Area	

For example, if a researcher is interested in 'informal settlement dwellers' in a city with many different informal settlements, a survey in a single settlement cannot be assumed to be representative of *all* informal settlement dwellers, given the differences in the people and ways of life across informal settlements (Auerbach, 2019). Residents of one informal settlement may be long-standing tenants with relatively secure status, while another informal settlement in the same city might be populated by recent migrants with tenuous claims to property rights or housing (Paller, 2019). In such a scenario, the survey can only claim to be representative of informal settlement dwellers *in a particular location.*

2.2 Where can they be found?

In some cases, the *location* and *type of place* where the population of interest can be found may not be obvious at the outset. The mobility of urban populations and their ability to 'hide' from surveyors can make compiling representative samples challenging. For example, urban migrants and urban refugees may seek to avoid identification out of concern for safety (Bloch, 2007; Vigneswaran, 2009) – this is not a problem just confined to LMICs. A study in London found that the 2011 census substantially undercounted the population of the

borough of Whitechapel due to a high concentration of undocumented migrants in the area (Flint Ashery, 2019). In such a situation, more resources and research may need to be devoted to the research design phase if a representative sample is sought (Vigneswaran, 2009). If the location of interest is large, the costs may become prohibitive and alternative research approaches need to be considered. Yet urban populations are typically not 'hidden,' but rather just hard to count (Farouk & Owusu, 2012). In other words, urban residents mostly do not actively resist being enumerated, though of course some sub-populations do. Yet in LMICs, many live in places that are difficult to access, and outsiders do not always feel comfortable going into these areas – or are not welcome e.g. in gated communities. Further, many of the neighbourhoods where these populations live are not 'on the map'-in officially designated areas that can be easily categorized and counted - and are ever-changing. Researchers have found different ways of overcoming this challenge, including hand-drawing maps that include registered traders and informal ones in Lagos, Nigeria (Grossman, 2020) and, similarly, workplace sampling of circular migrants in India (Thachil, 2018).

It is clear that identifying and locating populations requires significant and intensive field research before the data gathering process begins. Meeting respondents in their local 'habitat' or 'workplace' helps researchers achieve as representative sample of a population as possible. Researchers also need to consider when the population is present, which calls for a new approach beyond residence-based sampling.

2.3 What level of analysis?

Urban contexts display a remarkable heterogeneity with respect to geographic, social, and administrative boundaries. People even differ in their assessment of geography of their own communities or neighbourhoods (Burdick-Will, 2018; Coulton et al., 2001; Guest & Lee, 1984). Selecting the right boundaries and unit of analysis is important for inference, but also to inform policymakers due to challenges of multi-level governance (Kaufmann & Sidney, 2020). Some studies successfully overlay census tracts with electoral constituency boundaries, as they are often more salient to people's everyday lives (e.g. Nathan 2019).

Ultimately, one would like data and evidence to feed into policy, so 'for whom' the data is being collected is crucial. Organisations that fund data collection activities (such as IGC) often explicitly require government sponsorship at the outset. Even in the absence of explicit policy partnership, data collection, sampling and issues of representation should be addressed in light of the administrative units used by the government or policy implementation partner. For example, incidence of a disease for a geographical area that straddles two administrative units would be logistically difficult to respond to. Additionally, the response could be inefficient if the rate is

heavily influenced by one unit and not the other. While it is not always possible for sampling and administrative units to line up perfectly, it is necessary to think through the trade-offs and be as transparent as possible in the design phase.

Finally, ground-truthing - verifying approaches through information drawn from direct observation or measurement - helps to think through the who, where and what level of analysis questions; setting researchers up for a successful empirical strategy. There are different ways to do this, but many have relief on local experts to give feedback to approaches developed on paper. Yonto & Schuch (2020) ask local experts to provide feedback on pinpointing gentrification trends, enabling the researchers to 'map gentrification' at finer spatial resolution to the census tracts. Studies in India (Kit et al., 2012; Rains et al., 2019) and Ghana (Engstrom et al., 2015) use remote sensing to develop preliminary neighbourhood boundaries 'from the air,' but then consult with local stakeholders to refine the process.

Participatory mapping is another strategy used across many disciplines to give residents an active role and voice in the research process (Guldi, 2017; for a critical review see Panek, 2015). The Governance and Local Development Institute (GLD) has used focus group discussions with communities under study in Kenya, Malawi, and Zambia prior to face-to-face surveys to test out research questions, as well as confirm local boundaries. These studies highlight the utility of developing survey designs alongside qualitative research. We discuss this in more detail in the next section.

3 Sampling in cities without censuses

One of the biggest challenges for survey research in rapidly growing urban areas is achieving a representative sample of the population of interest. There are two primary hurdles to overcome: (i) devising a sample that is representative of an 'unknown' population, and (ii) accessing respondents. Here we focus on the first challenge; we address access in Section 5.

The standard approach to generating a representative sample is to begin with a complete list of individuals or households - usually from a census list - and randomly select those to be surveyed. If a census list is not available, alternative approaches can be used. Where the objective is to achieve a representative sample of a population living in a *particular area* (i.e. neighbourhood, district, city), spatial sampling is the best alternative. If the goal is to achieve a representative sample of a particular *sub-population* (e.g. refugees or other migrants), focusing on places where the sub-population may regularly be found (if not resident) has proven effective, as Thachil's study demonstrates (Thachil, 2018).

3.1 Sampling areas

Spatial sampling involves dividing an area into spatial units (often by overlaying a standardised grid over a map of the area of interest) and randomly selecting spatial units to survey. This approach can significantly improve coverage in urban areas with mobile or 'hidden' populations, yielding a more representative sample than the standard list-based approach if the list is out of date or partial (Landry & Shen, 2005). For example, in many cities, informal settlements are not officially counted, leaving them out of any official list that might be used to generate a sampling frame. Moreover, cities around the world exhibit high levels of racial, ethnic and socioeconomic clustering (Agyei-Mensah & Owusu, 2010; Iceland & Wilkes, 2006; Johnston et al., 2007), so a well-designed spatial sample can ensure representative coverage, even when demographic traits are not used to structure the sampling.

Significant improvements in computation, combined with the growing availability of areal imagery and user-friendly GIS software, have encouraged uptake of this approach for a wide range of applications over the past two decades. For example, spatial sampling techniques have recently been used to assess vaccination coverage in Niger

(Grais et al., 2007), respiratory health conditions in India (Kumar, 2007), dengue fever outbreaks in Nicaragua (Chang et al., 2009), mortality rates in Iraq (Galway et al., 2012), immigrant residential patterns in Johannesburg (Singh & Clark, 2013), and demographic estimates in Somalia (Driscoll & Lidow, 2014).

When sampling a relatively small area, such as a single neighborhood or informal settlement, randomly selecting grid cells and then households within those cells can be a cost-effective means of generating representative data. However, as the scale increases to the district or city-level, stratified sampling is required to ensure a representative sample can be obtained at reasonable cost (see Wang et al., 2012). This involves dividing the area of interest into a set of clusters from which to sample households or individuals. Generally speaking, these areas should conform to 'natural' boundaries in the urban landscape, which may or may not coincide with administrative boundaries (Hagen-Zanker et al., 2020).

There is no optimal way to draw 'natural' boundaries in urban areas, given that people often differ in their assessment of geography of their own communities or neighbourhoods (Burdick-Will, 2018; Coulton et al., 2001; Guest & Lee, 1984). Nevertheless, some have adopted a morphological approach using major transport infrastructural networks, such as roads, rails, canals, etc, to form urban residential zones (Grannis, 1998, 2005; Roberto & Korver-Glenn, 2021). These infrastructural features can be combined with other features of the urban landscape, such as critical slopes and large open lands to establish boundaries for clusters. Physical characteristics of buildings have also been used as proxies to identify spatial patterns and boundaries within settlements. Building density, size and roof type are among the features that have been explored so far (Baud et al., 2010; Kohli et al., 2012; Taubenböck & Kraff, 2014). There are freely accessible global datasets that can be used for this purpose, including Open Street Map (OSM) and Google Earth imagery.

Taking a morphological approach has strong merits because many people find built-environment barriers as salient features in the social and built landscape of cities. As Grannis (1998, 2005) shows, major infrastructure forms 'edges' in the built environment that assist people in coming to a common understanding of urban communities, but they *also* form nuclei of community. Further, the morphological cannot be understood in isolation: nearly 70% of respondents in Guest and Lee (1984) define their neighbourhoods according to *both* social and spatial criteria. Thus, it is critical to discuss the proposed areas with local experts and residents (ground-truthing, as discussed above), who can advise on which areas 'make sense' to locals to the area; this can then help the enumerator teams find their way around the selected areas (Hagen-Zanker, forthcoming).

There are useful examples of this in the literature, for instance Angeles et al. (2009) digitised existing maps of neighbourhoods in the six main urban areas of Bangladesh, then sent enumerators to verify and update these maps in a ground-truthing exercise and then re-processed the data to output a list of slum neighbourhoods. Without 'ground-truthing', 30% of the neighbourhoods would have been missed if they relied solely on remote sensing.

There have also been attempts to automate the process of boundary identification within cities. For example, Engstrom et al. (2015) extracted features from high-resolution satellite imagery to identify patterns in the urban landscape to map 'slum' settlements. Their results correspond very closely with ground-truthed maps provided by Accra Metropolitan Assembly and UN-Habitat. This approach could provide an accurate way of producing a slum inventory 'from the air.' More recently, advanced machine learning approaches such as convolutional neural networks have been applied in combination with satellite imagery to map deprived areas in Bangalore, India (J. Wang et al., 2019).

Spatial sampling approaches can also be used at the whole-city scale, although there are few examples of this to date. A notable one is Driscoll & Lidow (2014), who creatively estimated the population of Mogadishu, Somalia against a backdrop of acute insecurity and active fighting in the city. By combining areal imagery, a database of local contacts, and mobile technology, they used the square footage of (estimated) inhabitable spaces in the city as an initial proxy for the population, which they then updated based on images and in-person interviews. An additional technique used in Malawi is to crowd source geo-locations where residents live, ensuring that specific grids are not overlooked in the random walk procedure (Pasquale et al., 2017).

Once neighbourhoods and households have been sampled, a final step is to sample the respondent. For household surveys containing both household level and individual level questions, where respondents are sampled at their home, a final sampling choice has to be made on which household member to interview. Ignoring this step can lead to a biased sample which is not representative of the population and to biased survey estimates, if respondents' characteristics are related to constructs measured in the survey.

Moreover, non-response bias can result in samples that inaccurately represent the demographic characteristics of the studied population, even when efforts are made to correct this (Kreuter et al., 2010). In addition, a recent study in Karachi, Pakistan finds that respondents are systematically more likely to support ethnic politics and report greater feelings of perceived discrimination when in the presence of known others (Malik & Siddiqui, forthcoming).

Non-probability techniques, such as quotas based on demographic criteria like the youngest male in the household, or simply any adult from the household, are common. They are easy to implement, quick

and less intrusive, but, as already indicated, do not result in a representative sample of the target population (Smyth et al., 2019). As such, probability or quasi-probability methods of sampling, while more time-consuming to implement, lead to more representative data.

Random/ probability methods tend to be based on random number tables including the Kish grid (Kish, 1949) or random number stickers drawing on detailed household roster data (as used by the World Bank LSMS surveys).This can be time-consuming, especially in contexts where the majority of households are large (Eckman & Koch, 2019). Itcan also reduce response rates (Smyth et al., 2019) and may be inefficient if the roster data is otherwise not needed. Quasi-probability methods use shortcuts to sample respondents, for example interviewing the person with first or last birthday, or collecting more basic information on the household and respondent (Le et al., 2013). Such shortcuts are effective ways of reducing bias while keeping interview length manageable.

For a general overview of the theory and statistics associated with spatial sampling, see Wang et al. 2012. For a detailed, step-by-step guide to spatial sampling in human settlements see (Hagen-Zanker et al., 2020 and Hagen-Zanker, forthcoming).

3.2 Sampling sub-populations

Spatial sampling can also be used to generate representative samples of *subpopulations* rather than areas, especially those that are hidden or mobile. In this case, a purposive spatial sampling approach in the first stage may be required.

For example, a team of researchers based at the African Centre for Migration and Society document the socio-economic profiles of refugees living in Johannesburg (Vigneswaran, 2009). They started by purposively sampling three neighbourhoods where they knew refugee populations live. Then they randomly selected enumeration areas (EAs) using maps from Statistics South Africa, which each had approximately 300 dwellings. From each selected EA, 33 households were selected for interview. Density estimates were incorporated and EAs weighted by building size. This procedure ensured that populations typically overlooked in official surveys, like cross-border and internal migrants, were surveyed. Adida's 2014 survey of migrants in West African cities took a similar approach (Adida, 2014) while Landry & Shen (2005) used a similar technique in their survey of migrants in Beijing, relying on GPS locations from mobile phones and statistical techniques to estimate the size of migrant populations. The gridded population estimates enabled a representative sample for the subsequent individual attitude survey.

In some cases, an even more purposive spatial approach may be required in the form of time-location or space-time sampling. For example, (Agadjanian & Zotova, 2012) sought to survey 'irregular'

female labour migrants in Moscow from Kyrgystan, Tajikistan and Uzbekistan. Given the limitations of reaching this population through a traditional household survey, they use a workplace-based stratified probability sampling design. Targeting places of employment where the population of interest was known to be found proved to be a cost-effective effective and statistically more robust than respondent-driven sampling (ibid).

Respondent-driven sampling approaches, such as snowball sampling, are very common as they are easy and cost-efficient to apply but tend to result in a much higher statistical bias than other sampling methods. This bias can be reduced with a relatively simple two-stage design (Karon & Wejnert, 2012), such as those discussed above. Preliminary qualitative research can quickly reveal occupational or consumption-based clustering of subpopulations, which is common in cities, and this information can be used to design the sampling strategy. These insights suggest that creative strategies are needed to develop survey designs, which combine statistical approaches to demography and population estimates with spatial sampling approaches.

Finally, a topic that is rarely explicitly addressed is the challenge of reaching higher-income households and individuals (Inter-Secretariat Working Group on Household Surveys, 2022). In many contexts, this demographic lives in privatised spaces, e.g. gated communities, that are not accessible to enumeration teams. The value of their time may also preclude participation and response rates tend to be much lower (ibid). If capturing this hard-to-access demographic is needed to achieve a representative sample, a combination of time-location sampling and careful survey instrument design may improve the chances of successful data collection. It may also be necessary to recruit enumerators of the same ‘class’ that have a greater probability of being accepted in such spaces (such as university students from similar backgrounds).

4 Survey design

When designing the survey, it is important to place potential respondents at the centre of analysis (Wilson & Dickinson, 2021). While public attitude surveys like the *Afrobarometer* and *Latinbarometer* ask questions that are useful for formal politics and the state, and the *World Bank* social capital survey offers questions for community-level factors (Grootaert, 2004), there still is nogood model for understanding distinctly urban phenomena.

Cities and their neighbourhoods have unique characteristics that may require distinctive questions. Cities differ from rural areas with respect to daily practices, behaviours, and associational life (Weber, 1958; Wirth, 1938), spatial features like population density and built environment, underlying authority structures (Lust, 2022; Paller, 2019), modes of public order (Kushner, 2015; LeBas, 2013), and level of autonomy from the state (Tripp, 1997). Survey questions should consider with *whom* urban residents engage, as well as in *what* type of behaviour they are engaging.

In terms of content and length, once again we see that survey instruments for urban areas require adjustments, compared to traditional country-wide or rural populations. We now give some examples.

4.1 Types of survey items

Designing the survey instrument is an iterative process, which consists of two main phases. In the divergent phase, a comprehensive list of survey items is developed; then in the convergent phase these are narrowed down to the most critical survey items (Morra Imas & Rist, 2009). Individual survey items should be carefully considered in terms of their potential value to the analysis versus their cost including to overall survey length, respondent fatigue, respondent scepticism and repetitiveness. Other considerations include practical feasibility, sensitivity and relevance to local context. In urban settings, given the time constraints of respondents and access challenges, it is even more important to have a clear and short survey instrument.

Closed-ended surveys are efficient, straightforward, and common. They include dichotomous survey items, rating items (including Likert scales), categorical items and more (see Stantcheva, 2022 for a discussion of different options). Researchers should consider respondent experience to ensure the data collected is useful (Wilson

and Dickinson, 2022). Respondents often prefer questions that are easy to answer, such as dichotomous questions or Likert ranking questions with fewer options (Hagen-Zanker et al., 2022). Using too many different variations can be confusing for respondents and thereby increase interview length and fatigue, which has to be balanced with keeping the interviews interesting enough for the respondent.

Survey items should be formulated in simple, precise and neutral language, avoiding jargon, with a conversational tone to improve respondent experience, where possible (Hagen-Zanker et al., 2020; Stantcheva, 2022; Wilson & Dickinson, 2021). Each survey item should focus on one issue only, avoiding 'double-barrelled' survey items that ask about two outcomes at once (Morra Imas and Rist, 2009). In fact, breaking up questions into several easy to answer questions can be more effective and time saving, compared to more complex questions (Hagen-Zanker et al., 2020).

Once survey items and responses have been formulated, sequencing can be considered and instrument sections derived. Analytical themes do not necessarily need to go into the same instrument section, sometimes it can be more effective to break up certain themes over several sections to improve respondent experience (Hagen-Zanker et al., 2020). Sensitive issues are often better placed towards the end of the interview when rapport with the enumerator has been built. A clear and easy-to-understand survey is especially important in urban contexts, where the population is likely ethnically and linguistically diverse.

Stating the objectives of the survey is crucial, but can also be problematic. Respondents sometimes under-report assets if they think that government aid will be forthcoming after the survey (Martinelli & Parker, 2009). Explicitly stating the objectives of a survey can then raise expectations of aid and therefore affect the accuracy of some data. We find that neutral statements such, as 'understanding local conditions', which are true but do not go into great detail, work best. This is particularly true for more complicated survey designs such as treatment and control households across a spatial discontinuity, where the enumerators may not be skilled enough to understand the nuances of sampling, and where it is important that the control household not know about the intervention in the treatment area.

4.2 Survey length

Survey length is a key consideration in survey design. There is often a trade-off to be made between ambitions of the research, enthusiasm and curiosity of researchers versus the feasibility and quality of data collected. Longer surveys often result in lower quality data, as respondents and enumerators become more inattentive. For instance, respondents may be more inclined to acquiesce (or agree) with questions instead of answering honestly when fatigued towards

the end of the survey, or they may skip or not respond to questions (Smyth, 2016). This also has an ethical dimension: In most surveys, respondents are not being paid for time spent and many will be interrupted on a busy workday and, as such, shorter surveys are more respectful of respondents' time. This is particularly pertinent in an urban context, as respondents often have less time to spare with a more hectic lifestyle. This is especially relevant to workplace sampling protocols or street-intercept designs, which we discuss more below.

Extensive piloting and testing of the survey with members of the target populations is a critical step to collecting quality data. Piloting is needed to test the wording of survey items, length and structure of the survey, and the quality and usefulness of the data collected (Kelley et al., 2003). The pilot can also help refine options for close-ended, categorical questions to ensure only relevant (and all relevant) options are available (SLRC, 2013.). The pilot is also an opportunity to test the feasibility and practical application of the sampling approach and can highlight potential problems, such as poor response (Hagen-Zanker et al., 2020 and Kelley et al., 2003). Lastly, a pilot may highlight potential security issues or risks (SLRC, 2013).

5 Survey implementation

Successful surveys rely on a combination of factors, including: technical design, recruitment and training, ethical responsibilities, enumerator and respondent safety, quality control, and data storage. It is outside the scope of this paper to discuss all components of successful survey implementation. Instead, we focus on those that are distinct to rapidly growing urban areas, including technical innovations, recruitment, ethics, and public safety.

5.1 Technical innovations and data security

Computer-assisted personal interviewing (CAPI), which involves the use of electronic devices to administer surveys, is a rapidly emerging technique that holds much potential in improving the cost efficiency and quality of urban datasets (Benstead et al., 2017). Today, there are multiple freely accessible survey software that could be administered on mobile phones, tablets and other electronic devices. Some of these software, for instance the World Banks' Survey Solutions application, are fast gaining grounds in developing cities. Most of the software allow preloading of validation data against which responses can be cross-checked in real time and the pre-programming of various automated quality checks, thereby enhancing the quality of surveys. Data collected via CAPI platforms normally exist in formats compatible with major analytical statistical packages, helping to save time and avoid errors in data collation and digitisation.

As advanced earlier, CAPI techniques significantly reduce some of the inherent challenges addressed above with urban research. Most CAPI software not only come with preloaded maps, but also allow researchers to load their own base maps, which is particularly vital in capturing location data. For instance, in cases where a survey takes place away from a respondent's neighbourhood, the respondent could, with some assistance from an enumerator if needed, pin drop their location on a CAPI device. Location data captured on CAPI platforms can directly be exported to GIS platforms for further spatial analysis. This is particularly helpful in identifying starting points in hard-to-reach areas with no street addresses, as previous studies had to rely on hand-drawn maps and local knowledge that is often difficult to replicate (Paller 2019: appendix).

The use of tablets for data entry makes supervision and accountability much easier. Tablets can record the time taken for each survey, and deviations from a group average can indicate

outliers in quality. Since data is being entered in real time, this allows for efficient supervision as well. Often, data entry errors can be corrected at the request of a supervisor either on the basis of recall or on the basis of a revisit. In both circumstances, increasing the time lag between committing the error and the rectifying it can adversely affect the data quality and/or add to the survey cost. Explicitly budgeting for revisiting costs -recommended at 10% - is important to avoid rectifying errors on the basis of recall to the extent possible. Finally, CAPI software also records the length and time of interviews, providing another check for quality control.

5.2 Recruitment and training

Data collection requires the assistance of a team of data collectors, local resource persons, enumerators and their supervisors, who represent the research team to the respondent (United Nations, 2008). The method of hiring the enumeration team,either directly or through a survey firm, is equally important. Donor interest in collecting good quality quantitative and qualitative data over the last forty years has led to the emergence of enumeration as a profession, and many countries, particularly middle-income ones, now have well established survey firms. For example, the World Bank has conducted its flagship LSMS survey 150 times in more than 40 countries since 1985. Similarly, USAID implemented its DHS surveys in 90 low- and middle-income countries, and the UN's Multiple Indicator Cluster Survey has been carried out in 100 countries.

This being said, survey firms do not always specialise solely in policy or academic research and often conduct 'market' or opinion research on the side. This may become problematic since the level of rigour required for policy or academic research is very different to that required for market research. Understanding enumerator's contextual knowledge, as well as their biases, is crucial in assessing sources of measurement error (Jagger et al., 2011).

At least some of these errors can be anticipated and addressed by investing in rigorous and thorough training of enumerators and supervisors (Pullum et al., 2017). Preparing the training manuals for all persons who will be trained - interviewers, supervisors, and data entry staff (unless tablets are being used) - is perhaps the most important part of training (See Hagen-Zanker et al 2020 for a well-executed example).

Training manuals need to be written by the researchers, but updating the manuals on the basis of feedback from enumerators during training and pilot testing gives enumerators a sense of ownership that contributes to data quality (Jagger et al. 2012). In as much as the best-prepared manuals may have errors or omissions, additional instructions can be annexed during field work. In a recent survey conducted by one of the authors, simple polls implemented on a WhatsApp group of enumerators was found to be an effective tool for monitoring comprehension of additional instructions. Several days of

field-testing of the instrument - interviewing households followed by detailed feedback - is generally required before interviewers are ready to do their work effectively.

Finally, there is broad agreement that one should train more enumerators than required. This allows one to (a) test and select the best enumerators and (b) ensures a smooth workflow when enumerators (inevitably) drop out. This is especially true for surveys in urban areas, where in our own studies we find that enumerators often have more distractions, traffic problems, and accessibility challenges than in rural areas.

5.3 Informed consent and research ethics

Several ethical challenges emerge around informed consent that are particularly relevant in urban settings (Josephson & Smale, 2021). Firstly, populations may not be literate, and languages may not be written and/or 'written' consent might appear suspicious in the particular social context. In such cases, Institutional Review Boards have been known to allow oral consent.

Bhutta (2004) recommends audio recording of the consent - a procedure typically followed by qualitative researchers, and something easily achievable with tablets. Ultimately, the enumerator needs to ensure that respondents understand their rights within the research process. In one study, we had a successful experience of leaving copies of a one-page note (translated in the local and national languages) explaining the objectives of the study and leaving a contact phone number for complaints. While respondents themselves may or may not find the note comprehensible, the availability of the note and the possibility of cross-checking with a more knowledgeable neighbour or friend was found to be reassuring.

Secondly, respondents may be embedded in social structures that complicate the notion of 'individual' agency to give consent (Krogstad et al., 2010). Research protocols typically specify that individual respondents give consent. In rural areas, it is common to begin fieldwork by approaching a local leader or elder. In urban areas, it is quite possible that consent must be obtained at some collective level even if not in the centralised form as a tribal elder, yet the authority structure is more fluid and ambiguous. It is common to notify a group of concerned neighbours before an area can be surveyed. Sometimes, a collective consent is also crucial to ensure the safety and security of the enumerators, particularly female enumerators.

A third challenge, that is also apparent in some rural settings, is the requirement that a husband consent to the wife answering questions, particularly in his absence. In some cases, the absence of a husband (or his mother) may be unavoidable; it is recommended that this fact be recorded as part of the survey itself so one can at the very least control for it in econometric analysis. Ultimately, the respondent

needs to understand their rights from the research itself (Josephson and Smale, 2021; Krogstad et al. 2010; Bhutta 2004).

5.4 Safety and security

Considerations of safety are obvious for fieldworkers in conflict environments. But even in stable environments, one must not get caught off guard by a sudden deterioration of conditions. In our studies, one author almost got caught in crossfire in a place he did not consider particularly violent, while another enumerator had her purse stolen while carrying out an interview. Covid-19 has also brought to the forefront health and safety concerns.

It is always possible that someone might get hurt during fieldwork. Reconnaissance trips, identification of key informants, and environmental warnings provide initial risk assessments of the area. Our experience suggests three major concerns in urban-based field work that are not violence-prone or conflict zones: Sexual aggression, crowd hostility and health and well-being.

First, sexual aggression can affect both men and women. The pre-field work reconnaissance should elicit information on norms of sexuality, vulnerability and appropriate dress. Traveling in teams, dressing culturally appropriately, having a local emergency contact, and hiring available transport are key elements of a response. These underlying norms and risks may vary street-to-street and neighbourhood-to-neighbourhood, adding a layer of complexity to survey implementation in urban settings.

Second, large crowds have the potential to suddenly turn violent. Field work planning should ensure that there are no demonstrations, celebrations, or local holidays on fieldwork days that could affect the interview process.

Finally, issues of health and well-being are common. Water borne diseases like cholera and severe weather like rainstorms and flooding are common problems in urban settings. Enumerators and research teams need to be well-hydrated and suitably dressed for diverse weather possibilities.

5.5 Accessing respondents

Urban contexts pose specific challenges of accessing respondents. As discussed earlier, populations are often difficult to find. But there are also challenges when populations and locations are properly identified. Urban-based surveys still rely on random walk protocols based at identified starting points (Auerbach, 2019; Hagen-Zanker, forthcoming; Nathan, 2019; Paller, 2019), but the technological innovations addressed above do not overcome all accessibility challenges.

One problem is that certain subgroups of population are systematically more responsive to household surveys than others,

leading to sample estimates that are biased (Groves & Heeringa, 2006; Groves & Peytcheva, 2008). While few empirical papers report nonresponse rates, nonresponse biases remain an active area of research (Dutz et al., 2021).

The traditional approach to addressing unit non-response involves assuming that the non-response rate is random across relevant population subgroups and can therefore be removed by reweighting (see Berg, 2005 for example). But reweighting is based on the assumption that selection is on observables and cannot rectify the problem if the selection is actually on un-observables (Dutz et al, 2021).

In high-income countries, a number of surveys since Grooves and Heeringa (2006) have begun implementing “adaptive” designs, often based on information ancillary to the survey itself. These techniques focus more on non-response bias than the non-response itself. For example, incentives are often used to increase the response rate. But, as points out, even if incentives increase the response rate in general, they may change the composition of the sample. For instance, those that are persuaded by the incentive may be systematically different than those who are not. Monetary incentives, therefore, may be recommended under a response rate metric, but if one is interested in bringing in specific people, the survey needs to use a design to bring these people in, perhaps with a sacrifice of the non-response rate (see also Schouten et al., 2017 for a recent and exhaustive discussion). Adaptive design remains an active field. World Bank-sponsored surveys suggest stopgap solutions are needed to minimise respondent fatigue and maximise implementation quality.

6 Promising approaches to sampling and surveying urban populations in LMICs

Conducting surveys in urban settings is a complex and expensive process that involves substantial field research and experience. But it is also one of the only ways to gain a comprehensive picture of the social, political, and economic lives of urban dwellers. Based on the analysis above and the authors' 50 years of combined experience of conducting survey research in a wide range of contexts, we suggest a new approach to estimating urban populations that involves street-intercept short format questionnaires, based on intentional and partially administered survey modules. Researchers can then use post-statistical reconstruction of data to estimate the complete sample.

6.1 Street-intercept surveys

Street-intercept survey design is an alternative avenue to accessing populations traditionally uncounted, underrepresented and invisible in conventional household surveys. As we discuss above, researchers working in rapidly-growing environments that are data challenged normally face the hurdle of non-existent and quality-deficit spatial frame. Attempts at overcoming this mainly revolves around a gridded approach, or standardising a study area into grids and subsequently drawing sampling units (e.g. Driscoll and Lidow 2014; Landry 2010; Landry and Shen 2005). While the approach is innovative, it assumes the arbitrary drawn grids translate into meaningful urban morphological boundaries, an assumption that has weak grounding in urban geography theory.

The intercept survey technique can improve the quality of urban population data, and builds off the morphological approach. Buschmann defines it as "a method that allows to collect in-person survey responses in a public or private area from people who are in, or passing by, the same area at the time the data collection takes place" (Buschmann, 2019). Street-intercept surveys, therefore, involve sampling in spaces where the population is likely to be -in public spaces on the street, including: markets, transport stops, side-

of-road stalls, malls, and other busy areas. It ensures respondents engaged in different activities are sampled, including those working, running errands, travelling, and hanging out (Miller et al., 1997). The technique was first used to survey young women with HIV/AIDS (Nebot et al., 1994) and illicit drug users (Spooner et al., 1993).

The approach has several advantages. First, populations that prefer to keep their house of residence hidden, a demography that is rarely captured by household surveys, could still be reached via public and semi-public spaces. Conducting the survey in public places might also minimise distrust, especially in authoritarian contexts where home-based surveys might introduce feelings of discomfort (Buschmann 2019). Second, populations without house addresses or those homeless are accessible by street-intercept surveys.

Third, non-resident commuter populations excluded by household surveys are reachable in public spaces. Thus, the technique presents the opportunity for estimating the day time population of cities - residents and commuters - which is key for public service provision, infrastructure management, and other useful information. Fourth, the technique offers a much cheaper alternative to census-based and snowball-sampling procedures in order to estimate populations that are known to attend public areas or 'aggregation points,' with subsequent reweighting (McKenzie & Mistiaen, 2009).

However, whilst the technique generally improves access to hard-to-reach populations, it could also under-sample some sections of the population, particularly those who rarely use public and semi-public spacese.g. aging populations. In some contexts, like Mogadishu (Driscoll and Lidow, 2014), access to public spaces can also differ significantly by gender. This means potential systematic under-representations ought to be understood prior to the adoption of street-intercept survey design. Combining the technique with household surveys can overcome this challenge.

6.2 Short format questionnaires

Traditionally, population and demographic surveys take, at least, an hour to administer. Surveys normally consist of multiple modules and in most instances, respondents are required to answer all relevant ones, culminating into lengthy interviews. We suggest a technique that does not require respondents to answer all modules, but rather respond to a core module, and some randomised thematic modules. Shorter questionnaires have been effective in settings of fast-paced change, like protests, and offers the flexibility to capture spontaneous activity and fast-moving populations (Walgrave & Verhulst, 2011). They are also useful for sensitive questions, like political (Buschmann 2019) and health-related topics (Tourangeau & Yan, 2007)ebecause they allow significant time for the enumerator to discuss the items with the interviewee.

Researchers can then generate models based on statistical relationships between the core module and the partially assigned randomised thematic modules. The generated models can subsequently be used to reconstruct the data for unassigned or unanswered modules. One of the ways of validating the technique is to administer all modules to a sample of respondents and compare results. The results could also be cross-validated with existing secondary data, particularly in contexts with reliable datasets. By cutting down interview time, the technique both reduces survey cost, and enhances the appeal of street-intercept survey design (McKenzie and Mistiaen 2009).

6.3 Post statistical reconstruction

We suggest a novel machine learning technique called *sparse matrix reconstruction* that uses matrix factorisation to improve multiple imputation (Sengupta et al., 2023). This can be used to reconstruct a population after administering short format surveys. The technique has been successful at predicting missing values, and is now being used in the social sciences. The technique is especially useful for short format questionnaires that are administered to subsets of a population because it can handle sparse data sets and use out-of-sample testing or cross-validation for model selection (ibid: 74). It also advances the multiple imputation framework that is already standard practice in the social sciences (ibid).

The technique can help construct an as-if representative sample of urban populations because it is a two-dimensional regression model that simultaneously takes into account similarities across respondents, cases, or objects and similarities across responses, attributes, or features (Sengupta et al. 2022: 78). Therefore, it can take answers from the core module in a short format questionnaire, and predict answers for thematic modules using these imputation strategies. This could be especially useful for mobile and hard-to-find urban populations in cities, and significantly cut down the cost of face-to-face surveys. These three innovations developed alongside ground-truthing and engaged fieldwork would greatly enhance researchers' ability to collect data on hard-to-reach populations in the world's fastest-growing cities, contributing to our ability to achieve sustainable urban development.

7 Conclusions

Researchers, planners and policymakers in low- and middle-income countries often lack up-to-date and representative data about urban populations. Censuses are infrequent and expensive, while periodic surveys generally rely on sampling frames derived from censuses. In contexts of rapid population change, this can limit the quality and value of the data collected. Conducting surveys in urban settings is thus a complex and expensive process that involves substantial field research and experience. But it is also one of the only ways to gain a comprehensive picture of the social, political, and economic lives of urban dwellers. This is a high policy priority, as policies implemented in urban areas are critical for global efforts to tackle poverty and establish environmentally sustainable economic growth.

Based on more than 50 years of combined experience of conducting survey research in a wide range of contexts and a growing body of published research, we offer practical reflections and suggestions for dealing with some of the challenges of collecting data associated with doing surveys in urban contexts, with a particular focus on developing sampling approaches for urban areas without censuses.

We discussed issues related to research design, sampling in cities without reliable census data or lists, practical issues associated with the design of survey instruments and the process of enumeration. We have also put forward ideas for reducing the time and costs associated with gathering high-quality data in future survey exercises, including the use of street-intercept surveys, short-format survey instruments, and statistical reconstruction techniques.

Many of the considerations for carrying out surveys in urban areas are similar to standard surveys, for instance thinking about the ‘who, where and what level’ in terms of population, and keeping the survey instrument focused and targeted. Yet these considerations are even more pressing and critical in urban areas, given the sheer scale of urban populations, potential informality of neighbourhoods, employment and legal status and more limited time of urban dwellers. The most successful approaches have been innovative, combining different methodological approaches or types of data and including an element of ground-truthing -bringing in local expertise or data to verify or update approaches. As such, we end with a call for creativity and innovation, encouraging researchers to be inspired by the approaches developed so far, but to tailor approaches to their specific urban settings and populations to ensure that relevant data is collected.

References

Adida, C. L. (2014). *Immigrant Exclusion and Insecurity in Africa*. Cambridge University Press.

Agadjanian, V., & Zotova, N. (2012). Sampling and surveying hard-to-reach populations for demographic research: A study of female labor migrants in Moscow, Russia. *Demographic Research*, *26*, 131–150.

Agyei-Mensah, S., & Owusu, G. (2010). Segregated by neighbourhoods? A portrait of ethnic diversity in the neighbourhoods of the Accra Metropolitan Area, Ghana. *Population, Space and Place*, *16*(6), 499–516. https://doi.org/10.1002/psp.551

Angeles, G., Lance, P., Barden-O'Fallon, J., Islam, N., Mahbub, A., & Nazem, N. I. (2009). The 2005 census and mapping of slums in Bangladesh: Design, select results and application. *International Journal of Health Geographics*, *8*(1), 32. https://doi.org/10.1186/1476-072X-8-32

Auerbach, A. M. (2019). *Demanding Development: The Politics of Public Goods Provision in India's Urban Slums*. Cambridge University Press.

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan India: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation*, *12*(5), 359–374. https://doi.org/10.1016/j.jag.2010.04.008

Benstead, L. J., Kao, K., Landry, P. F., Lust, E. M., & Malouche, D. (2017). Using Tablet Computers to Implement Surveys in Challenging Environments. *Survey Practice*, *10*(2). https://doi.org/10.29115/SP-2017-0009

Berg, N. (2005). *Non-Response Bias* (SSRN Scholarly Paper No. 1691967). https://papers.ssrn.com/abstract=1691967

Bhutta, Z. A. (2004). Beyond informed consent. *Bulletin of the World Health Organization*, *82*, 771–777. https://doi.org/10.1590/S0042-96862004001000013

Bloch, A. (2007). Methodological Challenges for National and Multi-sited Comparative Survey Research. *Journal of Refugee Studies*, *20*(2), 230–247. https://doi.org/10.1093/jrs/fem002

Burdick-Will, J. (2018). School Location, Social Ties, and Perceived Neighborhood Boundaries. *City & Community*, *17*(2), 418–437. https://doi.org/10.1111/cico.12295

Buschmann, A. (2019). Conducting a Street-Intercept Survey in an Authoritarian Regime: The Case of Myanmar*. *Social Science Quarterly*, *100*(3), 857–868. https://doi.org/10.1111/ssqu.12611

Chang, A. Y., Parrales, M. E., Jimenez, J., Sobieszczyk, M. E., Hammer, S. M., Copenhaver, D. J., & Kulkarni, R. P. (2009). Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics*, *8*(1), 49. https://doi.org/10.1186/1476-072X-8-49

Coulton, C. J., Korbin, J., Chan, T., & Su, M. (2001). Mapping Residents' Perceptions of Neighborhood Boundaries: A Methodological Note. *American Journal of Community Psychology*, *29*(2), 371–383. https://doi.org/10.1023/A:1010303419034

Driscoll, J., & Lidow, N. (2014). Representative Surveys in Insecure Environments: A Case Study of Mogadishu, Somalia. *Journal of Survey Statistics and Methodology*, *2*(1), 78–95. https://doi.org/10.1093/jssam/smu001

Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A., & van Dijk, W. (2021). *Selection in Surveys: Using Randomized Incentives to Detect and Account for Nonresponse Bias* (Working Paper No. 29549). National Bureau of Economic Research. https://doi.org/10.3386/w29549

Eckman, S., & Koch, A. (2019). Interviewer Involvement in Sample Selection Shapes the Relationship Between Response Rates and Data Quality. *Public Opinion Quarterly*, *83*(2), 313–337. https://doi.org/10.1093/poq/nfz012

Engstrom, R., Sandborn, A., Yu, Q., Burgdorfer, J., Stow, D., Weeks, J., & Graesser, J. (2015). Mapping slums using spatial features in Accra, Ghana. *2015 Joint Urban Remote Sensing Event (JURSE)*, 1–4. https://doi.org/10.1109/JURSE.2015.7120494

Farouk, B. R., & Owusu, M. (2012). "If in doubt, count": The role of community-driven enumerations in blocking eviction in Old Fadama, Accra. *Environment and Urbanization*, *24*(1), 47–57. https://doi.org/10.1177/0956247811434478

Flint Ashery, S. (2019). *Micro-residential Dynamics: A Case Study of Whitechapel, London*. Springer International Publishing. https://doi.org/10.1007/978-3-030-00602-0

Galway, L., Bell, N., SAE, A. S., Hagopian, A., Burnham, G., Flaxman, A., Weiss, W. M., Rajaratnam, J., & Takaro, T. K. (2012). A two-stage cluster sampling method using gridded population data, a GIS, and Google EarthTM imagery in a population-based mortality survey in Iraq. *International Journal of Health Geographics*, *11*(1), 12. https://doi.org/10.1186/1476-072X-11-12

Grais, R. F., Rose, A. M., & Guthmann, J.-P. (2007). Don't spin the pen: Two alternative methods for second-stage sampling in

urban cluster surveys. *Emerging Themes in Epidemiology*, *4*(1), 8. https://doi.org/10.1186/1742-7622-4-8

Grannis, R. (1998). The Importance of Trivial Streets: Residential Streets and Residential Segregation. *American Journal of Sociology*, *103*(6), 1530–1564. https://doi.org/10.1086/231400

Grannis, R. (2005). T-Communities: Pedestrian Street Networks and Residential Segregation in Chicago, Los Angeles, and New York. *City & Community*, *4*(3), 295–321. https://doi.org/10.1111/j.1540-6040.2005.00118.x

Grootaert, C. (2004). *Measuring Social Capital: An Integrated Questionnaire*. World Bank Publications.

Grossman, S. (2020). The Politics of Order in Informal Markets: Evidence from Lagos. *World Politics*, *72*(1), 47–79. https://doi.org/10.1017/S0043887119000121

Groves, R. M., & Heeringa, S. G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *169*(3), 439–457.

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Guest, A. M., & Lee, B. A. (1984). How urbanites define their neighborhoods. *Population and Environment*, *7*(1), 32–56. https://doi.org/10.1007/BF01257471

Guldi, J. (2017). A History of the Participatory Map. *Public Culture*, *29*(1 (81)), 79–112. https://doi.org/10.1215/08992363-3644409

Hagen-Zanker, J. (forthcoming). Using satellite maps to build sampling frames in migration surveys. In W. L. Allen & C. Vargas-Silva (Eds.), *Handbook of Research Methods in Migration* (2nd ed.). Edward Elgar.

Hagen-Zanker, J., Hennessey, G., Carling, J., & Memon, R. (2020). Chapter 7: Survey data collection. In *MIGNIX Handbook*. Peace Research Institute Oslo. www.mignex.org/d031

Iceland, J., & Wilkes, R. (2006). Does Socioeconomic Status Matter? Race, Class, and Residential Segregation. *Social Problems*, *53*(2), 248–273. https://doi.org/10.1525/sp.2006.53.2.248

Jagger, P., Duchelle, A. E., Larsen, H. O., & Nielsen, O. J. (2011). Hiring, training and managing a field team. In *Measuring livelihoods and environmental dependence: Methods for research and fieldwork* (pp. 163–174). Earthscan. https://cgspace.cgiar.org/handle/10568/20885

Johnston, R., Poulsen, M., & Forrest, J. (2007). The Geography of Ethnic Residential Segregation: A Comparative Study of Five Countries. *Annals of the Association of American Geographers*, *97*(4), 713–738. https://doi.org/10.1111/j.1467-8306.2007.00579.x

Josephson, A., & Smale, M. (2021). What Do you Mean by “Informed Consent”? Ethics in Economic Development Research†.

Applied Economic Perspectives and Policy, *43*(4), 1305–1329. https://doi.org/10.1002/aepp.13112

Karon, J. M., & Wejnert, C. (2012). Statistical Methods for the Analysis of Time–Location Sampling Data. *Journal of Urban Health*, *89*(3), 565–586. https://doi.org/10.1007/s11524-012-9676-8

Kaufmann, D., & Sidney, M. (2020). Toward an Urban Policy Analysis: Incorporating Participation, Multilevel Governance, and "Seeing Like a City". *PS: Political Science & Politics*, *53*(1), 1–5. https://doi.org/10.1017/S1049096519001380

Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, *15*(3), 261–266.

Kish, L. (1949). A Procedure for Objective Respondent Selection within the Household. *Journal of the American Statistical Association*, *44*(247), 380–387. https://doi.org/10.1080/01621459.1949.10483314

Kit, O., Lüdeke, M., & Reckien, D. (2012). Texture-based identification of urban slums in Hyderabad, India using remote sensing data. *Applied Geography*, *32*(2), 660–667. https://doi.org/10.1016/j.apgeog.2011.07.016

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, *36*(2), 154–163. https://doi.org/10.1016/j.compenvurbsys.2011.11.001

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R. M., & Raghunathan, T. E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(2), 389–407. https://doi.org/10.1111/j.1467-985X.2009.00621.x

Krogstad, D. J., Diop, S., Diallo, A., Mzayek, F., Keating, J., Koita, O. A., & Touré, Y. T. (2010). Informed consent in international research: The rationale for different approaches. *The American Journal of Tropical Medicine and Hygiene*, *83*(4), 743–747. https://doi.org/10.4269/ajtmh.2010.10-0014

Kumar, N. (2007). Spatial Sampling Design for a Demographic and Health Survey. *Population Research and Policy Review*, *26*(5), 581–599. https://doi.org/10.1007/s11113-007-9044-7

Kushner, D. C. (2015). Nonstate Security and Political Participation: Reinforcing Ruling Party Support in South Africa. *Africa Today*, *62*(1), 107–135. https://doi.org/10.2979/africatoday.62.1.107

Landry, P. F., & Shen, M. (2005). Reaching Migrants in Survey Research: The Use of the Global Positioning System to Reduce Coverage Bias in China. *Political Analysis*, *13*(1), 1–22. https://doi.org/10.1093/pan/mpi001

Le, K. T., Brick, J. M., Diop, A., & Alemadi, D. (2013). Within-Household Sampling Conditioning on Household Size.

International Journal of Public Opinion Research, *25*(1), 108–118. https://doi.org/10.1093/ijpor/eds008

LeBas, A. (2013). Violence and Urban Order in Nairobi, Kenya and Lagos, Nigeria. *Studies in Comparative International Development*, *48*(3), 240–262. https://doi.org/10.1007/s12116-013-9134-y

Lust, E. M. (2022). *Everyday Choices: The Role of Competing Authorities and Social Institutions in Politics and Development*. Cambridge University Press.

Malik, M., & Siddiqui, N. (forthcoming). Third Party Presence and the Political Salience of Ethnicity in Survey Data. *Journal of Politics*.

Martinelli, C., & Parker, S. W. (2009). Deception and Misreporting in a Social Program. *Journal of the European Economic Association*, *7*(4), 886–908. https://doi.org/10.1162/JEEA.2009.7.4.886

McKenzie, D. J., & Mistiaen, J. (2009). Surveying migrant households: A comparison of census-based, snowball and intercept point surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(2), 339–360. https://doi.org/10.1111/j.1467-985X.2009.00584.x

Miller, K. W., Wilder, L. B., Stillman, F. A., & Becker, D. M. (1997). The feasibility of a street-intercept survey method in an African-American community. *American Journal of Public Health*, *87*(4), 655–658. https://doi.org/10.2105/AJPH.87.4.655

Morra Imas, L. G., & Rist, R. C. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank. https://doi.org/10.1596/978-0-8213-7891-5

Nathan, N. L. (2019). *Electoral Politics and Africa's Urban Transition: Class and Ethnicity in Ghana*. Cambridge University Press.

Nebot, M., Celentano, D. D., Burwell, L., Davis, A., Davis, M., Polacsek, M., & Santelli, J. (1994). AIDS and behavioural risk factors in women in inner city Baltimore: A comparison of telephone and face to face surveys. *Journal of Epidemiology & Community Health*, *48*(4), 412–418. https://doi.org/10.1136/jech.48.4.412

Paller, J. W. (2019). *Democracy in Ghana: Everyday Politics in Urban Africa*. Cambridge University Press.

Panek, J. (2015). How participatory mapping can drive community empowerment – a case study of Koffiekraal, South Africa. *South African Geographical Journal*, *97*(1), 18–30. https://doi.org/10.1080/03736245.2014.924866

Pasquale, A. D., McCann, R. S., & Maire, N. (2017). Assessing the population coverage of a health demographic surveillance system using satellite imagery and crowd-sourcing. *PLOS ONE*, *12*(8), e0183661. https://doi.org/10.1371/journal.pone.0183661

Pullum, T. W., Assaf, S., & Staveteig, S. (2017). *Comparisons of DHS estimates of fertility and mortality with other estimates*. ICF.

Rains, E., Krishna, A., & Wibbels, E. (2019). Combining satellite and survey data to study Indian slums: Evidence on the range of conditions and implications for urban policy. *Environment and Urbanization*, *31*(1), 267–292. https://doi.org/10.1177/0956247818798744

Roberto, E., & Korver-Glenn, E. (2021). The Spatial Structure and Local Experience of Residential Segregation. *Spatial Demography*, *9*(3), 277–307. https://doi.org/10.1007/s40980-021-00086-7

Schouten, B., Peytchev, A., & Wagner, J. (2017). *Adaptive Survey Design*. CRC Press.

Sengupta, N., Udell, M., Srebro, N., & Evans, J. (2023). Sparse Data Reconstruction, Missing Value and Multiple Imputation through Matrix Factorization. *Sociological Methodology*, *53*(1), 72–114. https://doi.org/10.1177/00811750221125799

Singh, G., & Clark, B. D. (2013). Creating a Frame: A Spatial Approach to Random Sampling of Immigrant Households in Inner City Johannesburg. *Journal of Refugee Studies*, *26*(1), 126–144. https://doi.org/10.1093/jrs/fes031

Smyth, J. D., Olson, K., & Stange, M. (2019). Within-Household Selection Methods: A Critical Review and Experimental Examination. In *Experimental Methods in Survey Research* (pp. 23–45). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119083771.ch2

Spooner, C. J., Flaherty, B. J., & Homel, P. J. (1993). Illicit drug use by young people in Sydney: Results of a street intercept survey. *Drug and Alcohol Review*, *12*(2), 159–168. https://doi.org/10.1080/09595239300185591

Stantcheva, S. (2022). *How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible* (Working Paper No. 30527). National Bureau of Economic Research. https://doi.org/10.3386/w30527

Taubenböck, H., & Kraff, N. J. (2014). The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data. *Journal of Housing and the Built Environment*, *29*(1), 15–38. https://doi.org/10.1007/s10901-013-9333-x

Thachil, T. (2018). Improving Surveys Through Ethnography: Insights from India's Urban Periphery. *Studies in Comparative International Development*, *53*(3), 281–299. https://doi.org/10.1007/s12116-018-9272-3

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

Tripp, A. M. (1997). *Changing the Rules: The Politics of Liberalization and the Urban Informal Economy in Tanzania*. University of California Press.

United Nations. (2008). *Designing Household Survery Samples: Practical Guidelines*. United Nations Statistical Division.

Vigneswaran, D. (2009). Residential Sampling and Johannesburg's Forced Migrants. *Journal of Refugee Studies*, *22*(4), 439–459. https://doi.org/10.1093/jrs/fep020

Walgrave, S., & Verhulst, J. (2011). Selection and Response Bias in Protest Surveys. *Mobilization: An International Quarterly*, *16*(2), 203–222. https://doi.org/10.17813/maiq.16.2.j475m8627u4u8177

Wang, J., Kuffer, M., Roy, D., & Pfeffer, K. (2019). Deprivation pockets through the lens of convolutional neural networks. *Remote Sensing of Environment*, *234*, 111448. https://doi.org/10.1016/j.rse.2019.111448

Wang, J.-F., Stein, A., Gao, B.-B., & Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics*, *2*, 1–14. https://doi.org/10.1016/j.spasta.2012.08.001

Weber, M. (1958). *The City*. The Free Press.

Wilson, L., & Dickinson, E. (2021). *Respondent Centred Surveys: Stop, Listen and Then Design*. SAGE.

Wirth, L. (1938). Urbanism as a Way of Life. *American Journal of Sociology*, *44*(1), 1–24. https://doi.org/10.1086/217913

Yonto, D., & Schuch, C. (2020). Developing and Ground-Truthing Multi-Scalar Approaches to Mapping Gentrification. *Papers in Applied Geography*, *6*(4), 352–368. https://doi.org/10.1080/23754931.2020.1789499