

Surveys and the City: Three Challenges to Quality Data Collection in Urban Areas

Millions of people move into cities in developing countries every month. What kind of a life awaits them there? The truth is: we don't really know. Development economists use household survey data to measure living standards across the world. Yet reliable data for cities in the global south is hard to come by. In this blog, I talk about three challenges to collecting data in cities – measurement, missing people, and money – as well as steps that can be taken in the design, implementation, and analysis of survey data to try and address them.

Measurement. The most basic challenge to understanding living standards in cities is that most household surveys are designed in such a way that they can tell us how the urban population differs from the rural population. To make inferences about differences between cities and within cities we would need to interview *many* more people. And that is expensive: on average, living standards surveys cost [170 USD per household interviewed](#). That means that it can cost hundreds of thousands of dollars to capture of representative sample for a city.

Small-area estimation (SAE) provides a partial solution to this problem. In simple terms, this technique uses statistical models to predict what responses *might* have been among households that did not participate in the survey. As long as we know a few key things about the households in question – for example, from a census – then we can try and predict outcomes like poverty rates based on the patterns that emerge from the survey data. SAE has many useful applications, but it relies on good models and reliable census data for the small area. This takes us to our second big challenge.

Missing people. The most vulnerable people in cities are often the hardest to survey, for both technical and political reasons. All surveys need a sampling frame – a list of the population from which you can draw the random sample to be surveyed. This is often the national census database. Yet it is not uncommon for census to be several years old. In contexts where cities are growing very quickly, large areas may be left out of the frame. In other contexts, they may be left out deliberately – for example, if it is politically expedient to ignore informal areas with high levels of migrant foreign workers.

Satellite imagery provides one way to help address this. It can be used as a check, and even as an alternative base from which to estimate population distributions. In the World Bank's [MLSC survey](#) – a survey of living standards piloted in Durban (South Africa) and Dar es Salaam (Tanzania) – satellite data was used to identify areas with more 'irregular' settlement patterns. This was then built into the sampling strategy so that data can be used to compare across visibly 'slum like' and 'non-slum like' areas.

The problems do not stop there, however. Even within areas, some people are likely to be excluded. For one, you need a fixed address: surveying homeless people and other transient populations requires entirely different sampling methods. Within the house, renters may be overlooked: in many developing countries tenants rent one or two rooms from an owner-occupier. Since these people do not eat with the owners, they are not counted as part of the household. Yet they also are not recorded as a separate household. In the MLSC

survey we built-in a step at listing to address this challenge, but there is room to explore further how wide-spread this challenge is and how best to address it.

Another issue is crime (and fear of crime). This makes people wary to talk to strangers; and perhaps rightly so, as criminals have been known to pose as interviewers to conduct robberies. On the other hand, interviewers may pretend that people refused to answer in order to spend less time in risky neighbourhoods. This can lead to biases in the data. Efforts to inform the local population ahead of the survey and to monitor of patterns in the data collected in real-time are key, but this is expensive and the methodology used is not always transparent. In practice, high non-response rates are largely addressed by adding weights to data analysis.

Money. Time is money in cities. Urban populations are busy and life is highly commoditised. This has implications for both what and how data is collected. Take poverty measures. All major international poverty lines – from the World Bank’s \$1.90 a day to the Multidimensional Poverty Index (MPI) – rely on household survey data. A number of people argue that they make urban poverty look less severe than it is. For example, if vital expenses such as transportation to work mean that households spend more money every day, people in cities may seem ‘richer’, but arguably are not better-off in practice. Costs of basic needs approaches don’t account for possible differences in the ‘bundle’ of goods needed to survive in urban versus rural areas. And even the MPI may suffer from limitations in [the choice of variables](#) used to capture poor living conditions (such as dirt floors, which may lead us to overlook unsafe conditions like illegal multistory tenement buildings).

The *methods* used to collect poverty data also matter. [Research shows](#) that small differences in the design of consumption modules – such as who answers the questions and the time-frame they are asked to report consumption for – can lead to dramatically different poverty headcount and inequality rates. People in cities tend to consume a wider range of goods, eat more food outside of the home, and face smoother market prices. This may make some methods more appropriate for cities than others.

Moreover, where life is highly monetised it is very important to get an accurate idea of prices. This is not straightforward. For example, housing is a huge expense for most people. How can we find out how much it costs? Surveys often ask homeowners to record the value *if, instead of owning your home, you had to rent it*. But people may not have a good idea of rental values. They may also overestimate the value of their own house simply because it belongs to them (in psychology this is called the ‘endowment effect’). In the MLSC survey we modified the question to: *If a friend of yours wanted to buy a property like this in the same neighborhood, how much would he/she have to pay?*

Surveys remain an important tool to understanding living conditions in urban areas, but there is considerable scope to refine approaches to collecting data in cities. After all, the world is already more than 50 percent urban and the total number of people living in cities is [set to double by 2050](#).

Alexandra Panman is a DPhil candidate in the Department of International Development, University of Oxford. She worked as a consultant in the team that developed the MLSC

surveys led by Nancy Lozano-Gracia as part of the broader World Bank Spatial Development of Cities Program.