 doi:10.20944/preprints201910.0320.v2

Feasibility of innovative tools and methods to improve household surveys in complex urban settings: Multiple methods analysis of the Surveys for Urban Equity (SUE) study in Kathmandu, Dhaka, and Hanoi

Dana R. Thomson,[1,*] Radheshyam Bhattarai,[2] Sudeepa Khanal,[2] Shraddha Manandhar,[2] Rajeev Dhungel,[2] Subash Gajurel,[2] Joseph Paul Hicks,[3] Duong Minh Duc,[4] Junnatul Ferdoush,[5] Tarana Ferdous,[6] Nushrat Jahan Urmy,[5] Riffat Ara Shawon,[5] Khuong Quynh Long,[4] Ak Narayan Poudel,[3] Chris Cartwright,[3] Hilary Wallace,[7] Tim Ensor,[3] Sushil Baral,[2] Saidur Mashreky,[5] Rumana Huque,[6] Hoang Van Minh,[4] Helen Elsey[3]

* Corresponding author: University of Southampton, Highfield Campus Building 58, Southampton SO17 1BJ, UK, Phone: +44(0)23 8059 2250, Fax: +44(0)23 8059 3131

1. Department of Social Statistics and Demography, University of Southampton, Southampton UK
2. Health Research and Social Development Forum-International, Kathmandu, Nepal
3. Nuffield Centre for International Health and Development, University of Leeds, Leeds UK
4. Hanoi University of Public Health, Hanoi, Vietnam
5. Centre for Injury Prevention and Research – Bangladesh, Dhaka Bangladesh
6. Advancement through Research and Knowledge Foundation, Dhaka Bangladesh
7. School of Medicine, The University of Notre Dame Australia, Freemantle WA Australia

ABSTRACT

Background: The methods used in low- and middle-income countries (LMICs) household surveys have not changed in four decades; however, LMIC societies have changed substantially. This mismatch may result in unintentional exclusion of vulnerable and mobile urban populations. We compare three survey method innovations with standard survey methods in Kathmandu, Dhaka, and Hanoi, and summarize feasibility of our innovative methods in terms of time, cost, skill requirements, and experiences.

Methods: We used descriptive statistics and regression techniques to compare respondent characteristics in samples drawn with innovative versus standard survey designs and household definitions, adjusting for sample probability weights and clustering. Feasibility of innovative methods was evaluated using a thematic framework analysis of focus group discussions with survey field staff, and via survey planner budgets.

Results: We found that a common household definition excluded single adult (46.9%) and migrant headed households (6.7%), as well as non-married (8.5%), unemployed (10.5%), disabled (9.3%), and studying (14.3%) adults. Further, standard two-stage sampling resulted in fewer single adult and non-family households than an innovative area-microcensus design; however, two-stage sampling resulted in more tent and shack dwellers. Our survey innovations provided good value for money and field staff experiences were neutral or positive. Staff recommended streamlining field tools and pairing technical and survey content experts during fieldwork.

Conclusions: This evidence of exclusion of vulnerable and mobile urban populations in LMIC household surveys is deeply concerning, and underscores the need to modernize survey methods and practices.

KEY WORDS

Nepal, Vietnam, Bangladesh, gridded population sampling, GridSample, OpenStreetMap, GeoODK, cross-sectional design, urban, household survey

 doi:10.20944/preprints201910.0320.v2

SUMMARY BOX

What is already known?

- Researchers and practitioners increasingly raise concerns about exclusion of vulnerable and mobile urban populations from LMIC household surveys on conceptual grounds: sample frames are usually outdated; typical two-stage designs require a long time gap between household listing (final sample frame) and interviews; and paper-based field tools and protocols developed 40 years ago are not well-suited to modern complex urban settings.
- LMIC urban settings pose numerous challenges to survey fieldwork including atypical housing arrangements, large numbers of migrant workers, rapid expansion of new – often informal – dwellings, and high mobility of residents.
- The challenges of conducting surveys in LMIC cities are only going to worsen as urbanization, population mobility, and socio-economic disparities increase, particularly in African and Asian cities.

What are the new findings?

- This study quantifies rates of exclusion among vulnerable and mobile sub-populations in Kathmandu Valley, and areas of Dhaka and Hanoi where these populations concentrate.
- We describe and evaluate innovative survey methods that might improve accuracy of household surveys in LMIC cities, including evaluation of feasibility.

What do the new findings imply?

- Alternative sample frames, such as gridded population estimates, are a viable alternative to outdated or inaccurate census sample frames. New types of sample frames can enable new survey designs, such as area-microcensus sampling, which improve coverage of vulnerable and mobile urban populations in surveys.
- Until urban areas can be stratified by deprived / not-deprived areas, or some other area classification that reflects urban disparities, household surveys are unlikely to accurately sample tent and shack dwellers in slum-like areas.

INTRODUCTION

In low- and middle-income countries (LMICs), household survey methods have remained consistent while population trends have changed substantially over forty years. This mismatch has likely increased exclusion of vulnerable and mobile populations from survey data. LMIC survey best-practices were established when LMICs were majority rural by agencies that have been critiqued for holding a “sedentary bias” in development initiatives.[1,2] Globally, human mobility has increased substantially over the last two decades, and today most LMICs are in the midst of urban transitions, or will be soon.[3] An estimated 2.5 billion people will be added to the planet by 2050, with 90% of that population increase concentrated in Asian and African cities alone.[4] While rates of urban growth in LMIC cities are consistent with rates previously observed in high income countries, the number of people added to LMIC cities today creates unprecedented scenarios of urbanisation. For example, Lagos Nigeria, Delhi India, and Dhaka Bangladesh are each expected to add more than 700,000 people per year through 2030.[4] Rapid in-migration to LMIC cities is accompanied by increased socio-economic inequalities, growth in slum populations, and housing crises, all of which contribute to increasingly complex living arrangements.[5,6] As urbanisation changes the structure and nature of communities and households in LMICs,[7] survey methods must evolve in response.

The largest survey programmes in LMICs include the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), and Living Standard Measurement Surveys (LSMS), which essentially use the same methods and tools.[8] Collectively, these programmes have performed nearly 700 national surveys in more than 130 countries since 1980. Across these surveys, census enumeration areas (EAs) are sampled with probability proportional to population size (PPS), households in selected EAs (i.e., clusters, primary sampling units) are mapped and listed, approximately 20 households are sampled in each cluster, and interviewers return later to administer questionnaires to selected households.[9–11] Among DHS surveys conducted since 2000, the average sample frame was seven years old (up to 30 years old), and 94% of surveys used the previous census as a sample frame, while the remaining 6% used an official list of areas or households.[12] By relying on census sample frames, unregistered and special populations excluded from the standard census are intentionally omitted from surveys including the homeless, internally displaced people, refugees, informal slum dwellers, nomadic populations, and institutional populations.[6,13]

Unintentional exclusion of vulnerable and mobile populations can additionally occur in in three ways. First, if structures built and occupied since the last census are systematically over-represented in deprived areas, vulnerable and mobile populations are systematically under-represented in the first-stage sample frame. Second, two-stage sample designs require a gap of several months between the mapping-listing and interview activities, resulting in systematic non-response from vulnerable and mobile populations not present at time of interview, and exclusion of recently occupied dwellings (living spaces). Third, disproportionate exclusion of vulnerable and mobile populations can result from poorly-defined or difficult to operationalize mapping-listing protocols; for example, assuming that one household occupies each dwelling. In this case, systematic under-listing of vulnerable and mobile households who share a dwelling results in their exclusion during the second stage of sampling.[14]

These three issues are labelled coverage error, non-response error, and sampling error, respectively, in the Total Survey Error framework, and threaten to bias survey results.[15] Additional measures of survey data relevance are of concern. Given the use of survey results by decision-makers to make inferences about the general population, intentional omission of the homeless, displaced populations, informal settlers, and others threatens relevance of survey results, particularly with respect to social and economic indicators.[15] Furthermore, without maps of deprived/non-deprived urban areas,[16] the survey results of the urban poorest are masked, or hidden, in aggregated urban averages resulting in limited relevance of survey results for decision-making.[15] Crucially, surveys are used to measure progress against one-fourth of the Sustainable Development Goal (SDG) indicators.[17] If current survey methods systematically under-represent and mask vulnerable and mobile urban populations, our understanding of progress towards the SDGs is fundamentally flawed.

To address problems of unintentional exclusion of vulnerable and mobile households in surveys, the Surveys for Urban Equity (SUE) project piloted and evaluated three survey innovations in Kathmandu, Dhaka and Hanoi: (1) use of modelled gridded population data as a sample frame which was assumed to be more current and have better coverage of the entire population than census, (2) area-microcensus sample design to remove the time-lag between mapping-listing and interviewing, and (3) mapper-lister protocols including a script, OpenStreetMap and OpenDataKit tools, and a broadened household definition to identify atypical dwellings and households. We were not able to obtain maps of deprived/non-deprived areas to stratify the surveys to address problems of robustness. Here, we present results of the pilot including the extent to which populations were unintentionally excluded from a standard survey design. Further, we evaluate the feasibility, cost and skills required to implement our novel methods in complex urban settings.

METHODS

We compared the ability of three survey innovations to identify different types of households and individuals than standard surveys. To establish feasibility of the innovations, we recorded costs and team skills required, and conducted focus group discussions (FGDs) to explore enumerator experiences.

Setting

We selected Kathmandu Nepal, Dhaka Bangladesh and Hanoi Vietnam, as they typify different points on the urbanisation trajectory. The pace of growth in South Asia has particularly strained urban housing markets, increasing the number of people living in atypical arrangements and locations.[3] While some poorer households live in informal settlements, others live in economically heterogeneous neighbourhoods.[3] In Kathmandu and Dhaka, for example, it is common for the building owner to occupy the top floor, rent the middle floor to a middle-class family, and rent the bottom floor to multiple low-wage workers. In Vietnam, old, cramped buildings continue to house the economically and socially vulnerable, while migrant labourers live in multiple-occupancy inadequate structures near work.[18] We sampled the entire Kathmandu Valley, and purposefully chose to survey a slum and an economically mixed ward in Dhaka, and an economically mixed district with a large migrant population in Hanoi. The Hanoi survey occurred soon after a government campaign to evict illegal occupants.

Study design and protocol

In 2017 and 2018, we conducted three cross-sectional household surveys in Kathmandu, Dhaka and Vietnam.[19]

 doi:10.20944/preprints201910.0320.v2

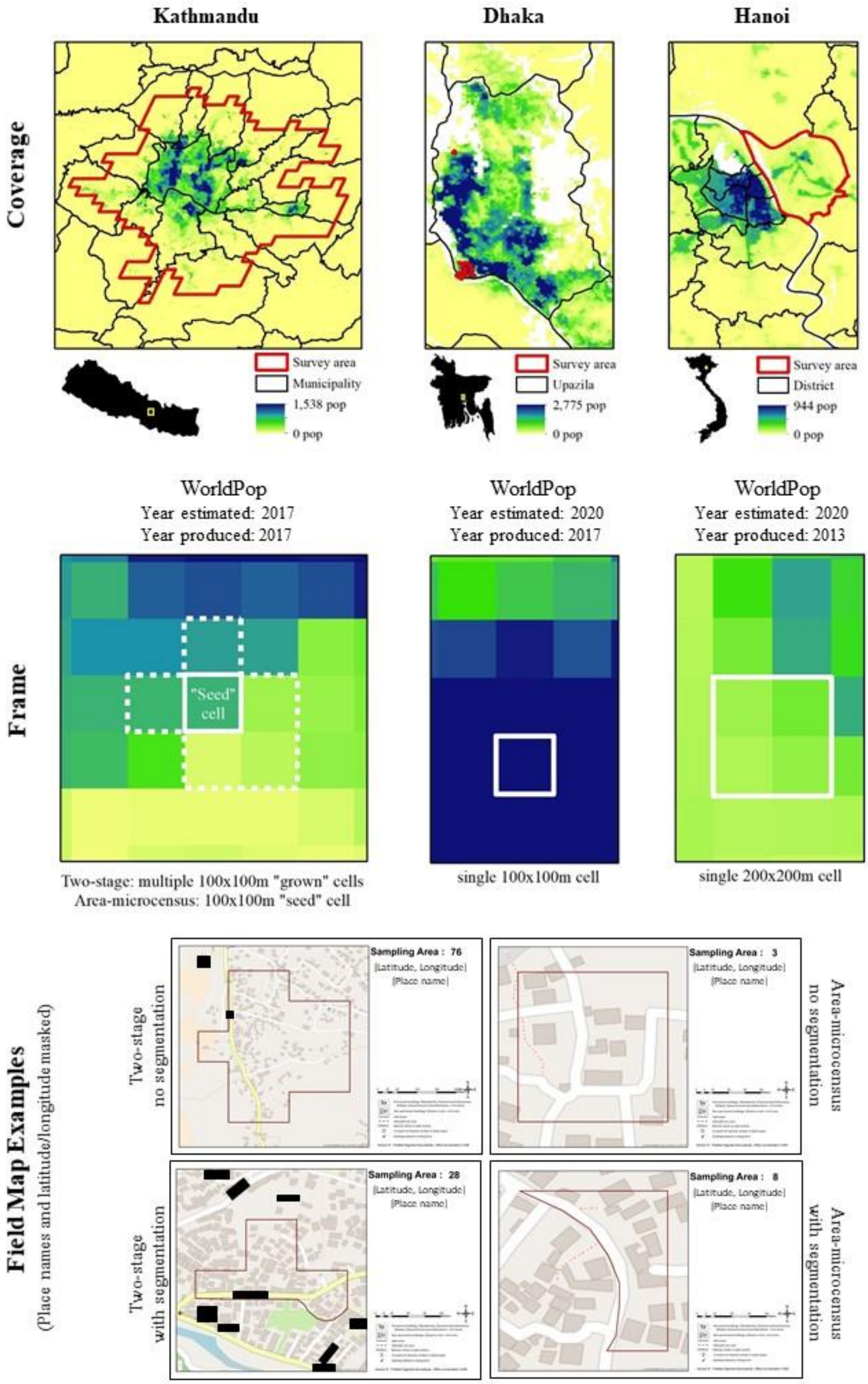

Figure 1. Surveys for Urban Equity coverage areas, sample frames, and example field maps in Kathmandu, Dhaka, and Hanoi

Coverage area. Nepal's government is in transition to a new federal republic system, and administrative boundaries were recently updated. Old Kathmandu municipality boundaries only included the city centre, while new municipality boundaries included rural communities beyond the peri-urban reach.[20] To ensure coverage of the functional city, we used the Global Human Settlement (GHS) layer of 1x1km grid cells defining "high dense urban" areas (Figure 1). In Dhaka, the survey covered one ward and one slum community, and in Hanoi, the survey covered one district (Figure 1).

Sample size. The survey in Kathmandu targeted 1200 households in 60 clusters to estimate depression and injury prevalence with a maximum margin of error of +/-4.27% with 95% confidence.[19] The Dhaka and Hanoi surveys targeted 400 households in 20 clusters each, with dual aims of evaluating transferability of SUE innovations whilst providing sufficient power (beta=10%) to estimate demographic and poverty indicators +/- 5% with 95% confidence.

Back-up clusters. Given the chance of selecting areas without residential buildings (e.g. airport or factory buildings) from gridded population data, and the possibility of selecting cells with no buildings, we selected 30% back-up clusters for each sample. This meant that we sampled 78 clusters in Nepal, and 26 clusters in Dhaka and Hanoi, before randomly assigning 60 (or 20) clusters to the main sample. If a sampled cluster had no residential buildings, then it was replaced with a randomly selected back-up cluster. Four additional back-up clusters were sampled in Hanoi after masking already selected clusters, because more than 6 clusters were dropped.

Sample design. Area-microcensus sampling means that all households in a cluster are sampled, allowing the household listing and interviews to occur on the same day. Area-microcensus sampling also allowed us to include populations typically omitted from surveys by design. In Kathmandu, we randomized half of the clusters to an area-microcensus arm and the other half to a two-stage arm to compare survey designs. In Dhaka, we used an area-microcensus design, stratified by ward/community with proportional allocation. The Hanoi survey followed an area-microcensus design, and was not stratified.

Sample frame. We used WorldPop gridded population estimates as sample frames rather than older censuses. At the time of planning, the last censuses in Nepal (2011), Bangladesh (2011) and Vietnam (2009) were seven or more years old.[21] WorldPop is modelled with a machine-learning approach that disaggregates population counts from administrative areas to approximately 100x100m grid cells based on dozens of spatial covariates derived from satellite imagery and GIS data.[22] The small size of grid cells enables area-microcensus sampling. The Kathmandu sample was drawn from 2017 WorldPop estimates, while the Dhaka and Hanoi surveys were drawn from 2020 WorldPop estimates produced in 2017, and 2013, respectively.[21]

Sample selection. At the time of survey, the GridSample R package was the only publicly available tool to perform PPS sampling from gridded population data.[23] The algorithm allows aggregation of population estimates to larger cells (e.g. 200x200m), and selection with PPS. Users can optionally "grow" non-overlapping clusters to a minimum population by randomly adding neighbouring cells to selected "seed" cells. This is not ideal, as sampling units should be formed before sampling; however, gridded population sampling tools with this capability were only recently developed.[24] We used the population in the "grown" sampling unit for sample weight calculations following the logic that a frame of "grown" sampling units is implied in the sample weights calculation (Supplement).[23] Theoretically an adaptive sample weight could be calculated;[25] however, the number of terms required for all combinations of potential cells that could be covered by the "growth" algorithm approaches infinity. In the Kathmandu two-stage sample, households were systematically sampled in Excel following standard methods.[9,10,26]

Cell size. In Kathmandu, all clusters were initially sampled from 100x100m cells and "grown" to a minimum of 820 people (approximately 200 households). Among these 60 selected clusters, half were randomized to the area-microcensus arm and given the boundary of the original 100x100m "seed" cell (Figure 1). In Dhaka, the sample frame comprised of 100x100m cells, and in Hanoi, the sample frame comprised of 200x200m cells (Figure 1). The optimum cell size for each survey was determined using satellite imagery (SUE training manual[26]).

Pre-field review and segmentation. We visualised each cluster boundary over satellite imagery in ArcGIS before producing field maps, and manually segmented clusters that clearly exceeded 200 (two-stage) or 20 (area-microcensus) households. Segment boundaries following roads and property fences, ensuring segments had approximately equal populations, then one segment was selected at random to represent the cluster (Figure 1).

Mapping-listing protocols. The mapping-listing trainings were each one-week and involved lectures, role-play, group discussion and a field test. Before fieldwork, mappers-listers updated buildings, roads, and pathways in each cluster in OpenStreetMap using the iDeditor tool.[27] In ArcGIS, the survey planning teams used the updated OpenStreetMap layer and cluster boundaries to create a geographically-accurate map for each cluster (Figure 1).[28] In the field, mappers-listers noted changes on the paper map, followed a script to approach residents, and upon request, distributed a written description of the survey. The household listing was collected in GeoODK, an OpenDataKit-based application,[29] for all buildings within the cluster or intersected by its boundary. Mappers-listers commuted from home to assigned nearby clusters using a provided stipend. Daily, they submitted listing records and an image of the field map, and periodically they visited the office to debrief and update OpenStreetMap with changes noted on paper maps.

Post-field segmentation (area-microcensus). To ensure that interviewers would find approximately 20 households in each area-microcensus cluster, any such cluster with more than 25 dwellings was segmented manually in ArcGIS by a GIS specialist and the survey coordinator after mapping-listing fieldwork, ensuring equal numbers of dwellings in each segment.[26]

Household definitions. The DHS and MICS define household members as: (i) usual residents or people who slept in the dwelling the previous night, who (ii) share living arrangements, and (iii) share food.[9,10] The LSMS defines household members as: (i) people who slept in the dwelling three or more of the last 12 months and (ii) share food.[11] By all DHS, MICS, and LSMS definitions, households in both residential and commercial buildings should be included,[9–11] guards and servants are subsumed into the household of their employment,[9–11] and seasonal and migrant populations are usually excluded by design.[30] The SUE household definition was broader and simply included all self-reported usual residents. The SUE definition additionally included hostel-dwellers and long-term occupants of guesthouses (defined as last 7+ consecutive days and working, looking for work, or in the city for another purpose such as supporting someone in hospital), and street-sleepers who slept in the cluster the previous night. Servants (and their families) who lived at the employer's residence were counted as a separate household.[26]

Interview protocols. In the Kathmandu two-stage arm, geospatial specialists mapped and listed households, while public health specialists conducted interviews with sampled households later. In Kathmandu and Dhaka's area-microcensus samples, geospatial experts mapped and listed *dwellings* and the household listing was performed by interviewers on the day of interview. Due to time constraints in Hanoi, mapping, listing, and interviews were wrapped into one activity and conducted by public health specialists. This meant that maps used by interviewers in Kathmandu and Dhaka were field-verified, while in Hanoi, maps had only been updated during pre-field enumeration using satellite imagery.

In all three surveys, the SUE household definition was used to determine eligibility, and respondents provided written informed consent, were 18+ years of age and usually a senior household member. The interviewers read questions and recorded responses on a tablet in GeoODK. The household questionnaire collected demographics, assets, income/savings/expenditures, social capital, migration, and injury information. We also collected information about living arrangements, meals, and length of time at the dwelling to classify individuals and households that met DHS/MICS and LSMS definitions during analysis. One adult in each household was randomly selected using the Kish method to complete an individual questionnaire with mental health and migration questions.[31]

Public involvement

Members of the public, including survey respondents, were not involved in setting the research questions, outcome measures, design, or implementation of the study, nor the dissemination of study results.

Statistical evaluation

Sample weights were calculated separately according to the SUE and DHS/MICS household definitions. In the area-microcensus samples in all cities, we evaluated whether use of the DHS/MICS household definition resulted in different estimates of individual and household characteristics compared to use of the SUE household definition using means or percentages, and linear or multinomial regression at 5% alpha level. In the Kathmandu sample, we used the same statistical techniques to compare whether characteristics differed in the area-microcensus versus two-stage sample; first, holding the DHS/MICS household definition constant, and second, comparing two-stage-DHS/MICS with area-microcensus-SUE households. Household characteristics included building type, member configuration, migration status of household head, slum household, and urban poverty index (UPI). [32] Individual characteristics included age-gender groups, employment status, marital status, and highest level of education. A reference group was selected for each variable to make statistical comparisons, and observations were dropped if they lacked data to determine household definition eligibility. We analysed survey results in Stata 14.0, adjusting for sample weights and clustering of observations. The analysis in Kathmandu was stratified by arm (area-microcensus/two-stage), and the analysis in Dhaka was stratified by community (ward/slum).

Days worked by each staff member and costs were recorded by the survey coordinator in each city. Time spent by survey coordinators to develop and learn the novel methods was excluded from cost calculations. However, time spent training mappers-listers and interviewers was included. In Kathmandu, we estimated costs for the area-microcensus and two-stage arms separately by holding constant costs of administration, training, and durable goods, and varying days of fieldwork.

Qualitative evaluation

An FGD was held with each of mapping-listing teams using the same guide covering topics of OpenStreetMap enumeration, mapping-listing, and workflow. Additional questions exploring differences in area-microcensus and two-stage clusters were included in the Kathmandu FGD. FGDs were facilitated and audio recorded by two trained qualitative researchers, and conducted in the local language. The recordings were transcribed into the local language and then translated into English. We performed a thematic Framework Analysis in NVivo 11, coding every line by theme and summarizing positive/neutral experiences, challenges, and recommendations.[33]

Ethics

Ethics approvals were obtained from the University of Leeds (ref:MREC16-137), University of Southampton (ref:26819), Nepal Health Research Council (ref:1761), Bangladesh Medical Research Council (ref:BMRC/NREC/RP/2016-2019/317), and Hanoi University of Public Health (ref:324/2017/YTCC-HD3).

RESULTS

In Kathmandu, 15% of clusters were dropped and replaced. No clusters were dropped in the targeted areas of Dhaka, and 45% were dropped and replaced in the Hanoi district (Table 1). Due to high density in Dhaka, and larger clusters in Hanoi, nearly all clusters in those cities required segmentation to achieve 20 households per cluster (Table 1). Household response rates were 96.8% in the Kathmandu two-stage arm, 88.3% in the Kathmandu area-microcensus arm, 98.7% in Dhaka, and 82.7% in Hanoi (Table 1). The treatment of survey arms as strata in the Kathmandu sample meant that weights were larger in the two-stage arm because clusters comprised larger populations (mean:1.673, range:0.298-5.524) than in the area-microcensus arm (mean:0.347, range:0.157-0.985) (Table 1). The root design effects (DEFTs) for key demographic and socioeconomic outcomes were larger in area-microcensus units for demographic indicators, but smaller in area-microcensus units for slum household, UPI, migrant status, and education indicators (Table 1).

Table 1. Number of clusters and households (unweighted), sample weights, and design effects by survey

	Kathmandu Two-stage		Kathmandu Area-microcensus		Dhaka Area-microcensus		Hanoi Area-microcensus	
Clusters								
Targeted	30		30		20		20	
Dropped and replaced	6		3		0		9	
Sampled	30		30		20		20	
Segmented	15		7		20		18	
Households								
Targeted	600		600		400		400	
Sampled - SUE	581		599		382		463	
Sampled - DHS/MICS (% of SUE definition)	578 (99%)		538 (90%)		318 (83%)		412 (89%)	
Sampled - LSMS (% of SUE definition)	578 (99%)		538 (90%)		343 (90%)		434 (94%)	
Household response rate	581/600 (96.8%)		599/678 (88.3%)		382/387 (98.7%)		463/560 (82.7%)	
Sample weights	*Mean (range)*		*Mean (range)*		*Mean (range)*		*Mean (range)*	
SUE	1.673 (0.298 - 5.524)		0.347 (0.157 - 0.985)		1.016 (0.113 - 2.595)		1.005 (0.196 - 4.123)	
DHS/MICS	1.581 (0.300 – 5.283)		0.346 (0.152 – 0.953)		1.012 (0.107 – 2.604)		0.931 (0.196 – 4.123)	
Design effects (SUE)	*Mean/prop. (SE)*	*DEFT*	*Mean/prop. (SE)*	*DEFT*	*Mean/prop. (SE)*	*DEFT*	*Mean/prop. (SE)*	*DEFT*
HH size	3.9 (0.111)	1.53	3.4 (0.137)	1.97	4.2 (0.178)	1.87	3.662 (0.110)	1.34
HHs per dwelling	1.0 (0.011)	2.11	1.9 (0.433)	4.20	2.2 (0.189)	2.68	Not recorded	--
HHs per PSU	19.5 (0.173)	4.42	24.9 (2.691)	5.40	20.9 (1.588)	4.96	34.6 (3.756)	6.05
Residential building	0.734 (0.023)	1.27	0.682 (0.075)	3.95	0.738 (0.065)	2.89	0.919 (0.020)	1.56
Nuclear family	0.517 (0.017)	0.83	0.439 (0.032)	1.56	0.535 (0.031)	1.20	0.500 (0.023)	0.96
Slum household	0.217 (0.452)	2.43	0.172 (0.33)	2.13	0.330 (0.044)	1.83	0.919 (0.023)	1.84
Slum household (without tenure)	0.184 (0.039)	2.39	0.140 (0.031)	2.18	0.275 (0.043)	1.87	0.008 (0.006)	1.38
Urban poverty index	0.320 (0.060)	3.08	0.229 (0.038)	2.21	0.770 (0.032)	1.50	0.040 (0.019)	2.11
Migrant (head of HH)	0.700 (0.056)	2.96	0.780 (0.025)	1.48	0.543 (0.034)	1.32	0.665 (0.070)	3.22
Married	0.675 (0.014)	1.23	0.663 (0.026)	2.13	0.758 (0.017)	1.30	0.723 (0.018)	1.46
Employed full-time	0.459 (0.022)	1.82	0.486 (0.028)	2.21	0.523 (0.019)	1.20	0.584 (0.034)	2.47
Male 18+	0.371 (0.013)	1.34	0.416 (0.022)	2.02	0.319 (0.009)	0.79	0.317 (0.017)	1.52
Secondary+ education	0.495 (0.042)	3.99	0.528 (0.032)	2.95	0.145 (0.014)	1.54	0.568 (0.014)	1.13

Unintentional exclusion due to household definition

Across the area-microcensus samples, applying the DHS/MICS or LSMS household definition resulted in exclusion of approximately 10% of households (unweighted) compared to the SUE definition (Table 1). In Kathmandu, nearly half (46.9%) of single adult households and sizable portions of migrant-headed households (6.7%), non-married (8.5%), unemployed (10.5%), disabled (9.3%), and studying (14.3%) adults were excluded by the DHS/MICS definition (Table 2). In the Dhaka and Hanoi surveys targeting vulnerable communities, sizable portions of single adult households (95.0% and 47.6%), non-married (48.1% and 37.3%), unemployed (32.6% and 23.9%), retired (70.5% and 27.6%), disabled (48.9% and 55.2%), studying adults (81.4% and 84.0%), young people (59.4-79.8% and 88.5-92.7%), and adult women (50.6% and 18.4%) were excluded by the DHS/MICS household definition (Table 2)

Table 2. Unintentional exclusion due to household definition: Percent of population who would be excluded using the standard DHS/MICS versus SUE household definition in Kathmandu, Dhaka, and Hanoi

Households in each area-microcensus sample were split by those who (a) met the SUE and DHS/MICS household definitions, and (b) met the DHS/MICS household definition only. We present the percent of households excluded from the DHS/MICS household definition, and regression coefficient p-value comparing (a) and (b).

Indicator	**Kathmandu Area-microcensus sample only**				**Dhaka Area-microcensus sample**				**Hanoi Area-microcensus sample**			
	N-wgt all	N-wgt DHS/MICS only	% excluded by DHS/ MICS	p-value†	N-wgt all	N-wgt DHS/MICS only	% excluded by DHS/ MICS	p-value†	N-wgt all	N-wgt DHS/MICS only	% excluded by DHS/ MICS	p-value†
Households												
Configuration												
Single adult	22	12	46.9	<0.001	24	1	95.0	<0.001	43	23	47.6	0.002
One woman with children	10	10	0.0	<0.001	9	8	7.9	0.967	6	2	66.7	0.006
Nuclear family	91	91	0.6	Ref.	205	188	8.3	Ref.	231	228	1.4	Ref.
Other family *	73	73	0.6	0.906	143	128	10.6	0.579	147	136	7.0	0.042
Non-family	13	13	0.0	<0.001	1	0	89.5	0.013	35	20	42.6	0.001
Slum household ** (with security of tenure)												
No	172	164	5.1	Ref.	295	248	15.9	Ref.	31	25	17.7	Ref.
Yes	36	33	6.6	0.809	87	77	11.4	0.281	425	382	10.1	0.494
Missing	0	0	--	--	0	0	--	--	7	2	72.3	0.120
Slum household ** (without security of tenure)												
No	179	170	5.0	Ref.	318	268	15.8	Ref.	456	404	11.4	Ref.
Yes	29	27	7.4	0.722	64	57	10.3	0.341	4	3	7.4	0.711
Missing	0	0	--	--	0	0	--	--	3	2	31.1	0.112
Urban poverty index												
Non-poor	161	152	5.2	Ref.	88	79	9.7	Ref.	444	396	10.9	Ref.
Poor	48	45	5.7	0.930	294	246	16.5	0.164	19	14	23.4	0.160
Migration status (head)												
Non-migrant	46	46	0.3	Ref.	174	156	10.6	Ref.	155	140	10.0	Ref.
Migrant	162	151	6.7	0.016	208	169	18.5	0.170	308	270	12.1	0.483
Adults 18+												
Marital status												
Not married	184	169	8.5	0.001	247	128	48.1	<0.001	331	208	37.3	0.001
Married	364	355	2.3	Ref.	779	548	29.6	Ref.	868	794	8.6	Ref.
Missing	0	0	--	--	1	1	0.0	<0.001	3	2	32.0	0.310
Employment status												
Full-time employed	267	262	1.6	Ref.	538	493	8.3	Ref.	702	653	6.9	Ref.
Part-time, underemployed	10	10	0.0	<0.001	37	32	12.5	0.556	39	37	7.0	0.989
Unemployed	27	24	10.5	0.001	46	31	32.6	0.003	92	70	23.9	0.007
Retired	20	19	1.9	0.839	307	91	70.5	<0.001	46	33	27.6	0.041

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 April 2020 doi:10.20944/preprints201910.0320.v2

Homemaker	123	122	1.5	0.860	2	1	46.6	0.133	215	184	14.4	0.004
Disabled "unable to work"	17	16	9.3	0.009	34	18	48.9	0.002	21	9	55.2	<0.001
Student	82	70	14.3	0.003	57	11	81.4	<0.001	82	13	84.0	<0.001
Missing	2	0	100.0	<0.001	6	2	75.2	0.012	5	4	19.0	0.448
Individuals												
Gender and age group												
Male <12	55	54	1.4	0.139	206	47	77.3	<0.001	207	22	89.6	<0.001
Female <12	48	47	1.6	0.291	180	36	79.8	<0.001	157	18	88.5	<0.001
Male 12-17	31	30	4.9	0.822	105	42	59.8	<0.001	78	6	92.7	<0.001
Female 12-17	32	31	3.4	0.442	87	35	59.4	<0.001	47	4	90.7	<0.001
Male 18+	297	280	5.7	Ref.	512	422	17.5	Ref.	536	460	14.2	Ref.
Female 18+	251	244	2.8	0.203	514	254	50.6	<0.001	665	543	18.4	<0.001
Missing	0	0	--	--	2	2	0.0	<0.001	0	0	--	--
Level of education												
Less than primary	171	163	4.7	0.733	906	443	51.2	0.062	340	69	79.8	<0.001
Primary	124	118	4.6	0.711	353	195	44.9	0.803	232	149	36.0	0.012
Secondary+	377	362	3.9	Ref.	233	131	43.6	Ref.	960	813	15.4	Ref.
Missing	42	42	0.0	<0.001	113	70	38.1	0.449	158	23	85.5	<0.001

* includes living with servants and/or extended family, sometimes with non-family household members as well
** defined as lacking improved water, improved sanitation, a durable structure, sufficient sleeping space (based on DHS/MICS household member definition), or insecure tenure
† multinomial logistic regression
N-wgt – weighted count

Unintentional exclusion due to sample design

Applying the DHS/MICS household definition, we compare area-microcensus and two-stage samples in Kathmandu to understand how sample design might influence types of respondents (Table 3). We found average household size was smaller in the area-microcensus sample but dwellings had more occupants (household: 3.5 vs. 3.9, dwelling: 5.0 vs. 3.9) (Table 3). Further, the area-microcensus design had more non-family households (6.0% vs. 1.9%), but the two-stage design included more shack and tent dwellers (0.7% vs. 3.8%) (Table 3).

Table 3. Unintentional exclusion due to sample design and household definition: Kathmandu sample characteristics comparing a) two-stage DHS/MICS versus area-microcensus DHS/MICS, and b) two-stage DHS/MICS versus area-microcensus SUE

Indicators	Two-stage DHS/MICS (Ref.)		Area-microcensus DHS/MICS			Area-microcensus SUE		
	N-wgt	Mean or Percent	N-wgt	Mean or Percent	p-value†	N-wgt	Mean or Percent	p-value†
Survey Metrics								
HH size	928	3.9	191	3.5	0.014	208	3.4	0.013
Dwelling size	928	3.9	191	5.0	<0.001	208	5.3	0.001
HHs per PSU	928	19.5	191	23.4	0.016	208	24.9	0.051
Households								
Building Type								
Residential	681	73.4 %	137	71.8 %	Ref.	142	68.2 %	Ref.
Mixed	206	22.2 %	50	26.4 %	0.595	52	25.0 %	0.594
Commercial	6	0.7 %	3	1.2 %	0.447	2	1.2 %	0.450
Shack or tent	35	3.8 %	1	0.7 %	0.009	1	0.6 %	0.009
Hostel	0	--	0	--	--	8	3.8 %	<0.001
Street-sleeper	0	--	0	--	--	2	1.0 %	<0.001
Guesthouse	0	--	0	--	--	0	0.1 %	<0.001
Configuration								
Single adult	42	4.5 %	11	5.8 %	0.256	22	10.4 %	0.040
One woman with children	29	3.2 %	10	4.9 %	0.093	10	4.7 %	0.096
Nuclear family	480	51.7 %	88	46.1 %	Ref.	91	43.9 %	Ref.
Other family*	360	38.8 %	70	36.8 %	0.600	73	35.1 %	0.603
Non-family	17	1.9%	12	6.3%	0.029	13	6.0%	0.030
Slum household** (with tenure)								
No	729	78,5%	158	83.0 %	Ref.	172	82.8 %	Ref.
Yes	199	21,5%	32	17.0 %	0.393	36	17.2 %	0.418
Urban poverty index								
Non-poor	633	68.2 %	147	77.2 %	Ref.	161	77.1 %	Ref.
Poor	295	31.8 %	44	22.8 %	0.189	48	22.9 %	0.201
Migrant (Head)								
No	280	30.1 %	44	23.2 %	Ref.	46	22.1 %	Ref.
Yes	648	69.9 %	147	76.8 %	0.244	162	78.0 %	0.173
Adults 18+								
Marital status								
Not married	861	32.5 %	163	32.2 %	0.924	185	33.7 %	0.107
Married	1,786	67.5 %	344	67.8 %	Ref.	363	66.3 %	Ref.
Employed full-time								
No	1,430	54.0 %	253	49.9 %	0.253	280	51.1 %	0.430
Yes	1,217	46.0 %	254	50.1 %	Ref.	267	48.7 %	Ref.
Missing	0	--	0	--	--	1	0.3 %	<0.001

continued…

... continued

Individuals								
Age, Gender group								
Male <12	334	9.4 %	52	7.9 %	0.149	55	7.7 %	0.089
Female <12	232	6.5 %	46	6.7 %	0.875	48	6.7 %	0.710
Male 12-17	170	4.8 %	29	4.3 %	0.287	31	4.4 %	0.275
Female 12-17	181	5.1 %	30	4.5 %	0.330	32	4.5 %	0.275
Male 18+	1,329	37.3 %	271	40.8 %	Ref.	297	41.6 %	Ref.
Female 18+	1,318	37.0 %	236	35.6 %	0.202	251	35.2 %	0.118
Education								
Less than primary	957	26.9 %	157	23.8 %	0.412	171	23.9 %	0.440
Primary	599	16.8 %	115	17.3 %	0.880	124	17.4 %	0.906
Secondary+	1,774	49.8 %	351	52.9 %	Ref.	377	52.8 %	Ref.
Missing	234	6.6 %	41	6.1 %	0.601	42	5.9 %	0.494

* includes living with servants and/or extended family, sometimes with non-family household members as well
** defined as lacking improved water, improved sanitation, a durable structure, sufficient sleeping space, or insecure tenure
† linear regression coefficient (continuous) or multinomial logistic regression (categorical)
N-wgt – weighted count

Unintentional exclusion due to sample design and household definition

Building off the previous analysis, we compared the area-microcensus sample with SUE definition and the two-stage sample with DHS/MICS definition in Kathmandu to understand the combined effects of survey design and household definition. In the area-microcensus-SUE sample, there were more single adult (10.4% vs. 4.5%) and non-family households (6.0% vs. 1.9%), plus inclusion of hostel-dwellers (3.8%), street-sleepers (1.0%), and long-term guesthouse residents (0.1%) who did not meet the DHS/MICS household definition (Table 3). However, the two-stage-DHS/MICS sample included more shack and tent dwellers (0.6% vs. 3.8%) (Table 3).

Time and cost

In Kathmandu, the area-microcensus gridded population survey arm with a target of 600 households in 30 clusters cost approximately US$26,769, or US$45 per household, while a comparable two-stage survey cost approximately US$35,284, or US$59 per household. Area-microcensus survey costs per household in Dhaka (US$34) and Hanoi (US$76) differed due to cost of living and limited economy of scale in those smaller samples. The main cost difference between Kathmandu's survey arms was the mapping-listing activity; costs were 2.5 times greater in the two-stage arm due to larger clusters.

Skill mix

The skills required to plan and implement SUE surveys were similar to standard household surveys. The main difference was skillset of the mapping-listing team. In a standard survey, mapping-listing staff are required to have a secondary education.[34] To use SUE tools and methods, the mapping-listing staff should additionally have training in geography, GIS, or related fieldwork, and be comfortable using mobile technologies for data collection and navigation. The skillsets of other staff including survey planners, trainers, and interviewers were identical to a standard household survey. The GridSample R package required intermediate R programming and GIS skills; however, a free point-and-click tool called gridsample.org is now available, allowing non-technical design and implementation of gridded population surveys.

Table 4. Comparison of time and budget to perform area-microcensus versus two-stage survey (estimated) in Kathmandu, Dhaka, and Hanoi

Budget Item	**Kathmandu, Two-stage**		**Kathmandu, Area-microcensus**		**Dhaka, Area-microcensus**		**Hanoi, Area-microcensus**	
	Time	Cost USD	Time	Cost USD	Time	Cost USD	Time	Cost USD
Planning & Administration	75 days		60 days		60 days		20 days	
Salaries		9,240		8,006		4,305		7468
Mapping-Dwelling/HH listing-GIS	35 days × 6 mapper-listers 1 GIS specialist		12 days × 6 mapper-listers, 1 GIS specialist		36 days × 8 mapper-listers, 1 GIS specialist		8 days × 12 listers	
Salaries, per diem		7,641		3,056		4,926		6128
Materials		291		218		120		68
Interviews & Data Management	19 days × 8 interviewers		15 days × 8 interviewers		24 days × 7 interviewers		13 days × 12 interviewers	
Salaries, per diem		5,723		4,518		2,345		11,872
Materials, including pilot		2,106		2,106		872		574
Incentives, local collaborators		0		0		0		3,089
Ethics review		1,998		1,998		238		1,362
Equipment								
Laptops / hard drives		1,193		1,193		167		0
Tablets *		1,212		1,212		382		1,714
Overhead	(20% direct costs)	5,786	(20% direct costs)	4,367	(20% direct costs)	2,671	(10% direct costs)	3,228
TOTAL		35,284		26,769		16,026		35,503
Per household		**59**		**45**		**34**		**76**

Experiences

Feedback from the mapper-lister FGDs was generally neutral or positive, and staff resoundingly said they would prefer SUE tools and protocols to a conventional paper-based protocol. The SUE survey fieldwork, however, was not without limitations.

Key challenges. In Kathmandu, the mapping-listing staff were comprised of university geospatial students. Several described approaching residents as their greatest challenge, as well as their greatest reward. One mapper-lister explained, "It was fun to work at the social level and interacting with the local people. We always used to be limited to using the computers before." Mappers-listers added that role-play and practical activities prepared them for fieldwork, though additional training on the survey aims would have helped to explain the survey's purpose to residents. In Kathmandu, mapping-listing staff initially enumerate 20-30 households daily, and this increased to 40-50 households daily after a week.

The challenges in Dhaka and Hanoi were different. In these cities, the survey planners were trained about SUE tools and protocols but did not have field experience before training mapper-listers and interviewers. As a result, mapping-listing staff, including the geospatial students in Dhaka, described challenges using the tablet applications during the first days of fieldwork. In Hanoi where public health experts performed mapping, listing, and interviews, staff additionally struggled with navigation. Due to community scepticism following recent government evictions in Hanoi, teams enlisted local guides to help approach residents and introduce the survey.

Across cities, mappers-listers described working in pairs as essential because it provided them with "mutual support" to adapt to the moods and reactions of residents, interact in more languages, and to work faster with more accuracy. Overwhelmingly, mappers-listers recommend that teams be comprised of one geospatial and one public health specialist.

Response rates. In all three cities, mapping-listing staff reported that residents seemed to omit mention of neighbours who did not have official mortgages or rental contracts, presumably for fear of evictions or fines. This was a particular challenge in Hanoi where "people tended to answer our question following their household record book," an official registry of households administered by the government. One mapper-lister-interviewer explained, "for residents who were living in evacuated houses, they felt worry and scare as if something wrong could happen."

In Hanoi, teams returned to each cluster multiple times to build trust with residents and identify households not reported during previous visits. While the presence of guides likely improved response rates, it also meant that survey teams were limited by guides' schedules. Most teams performed the listing and interviews in the evenings when guides were home, though this meant that residents were eating dinner and rushed, or refused. Mapper-listers and interviewer in Kathmandu and Dhaka performed their work during the day.

Residential building access was a problem across cities. The Hanoi teams faced secured apartment buildings without a guard. In these situations, the planning team contacted the building management boards and were usually able to gain access to these buildings, however once inside, mappers-listers-interviewers often found that residents knew little about their absent neighbours. Kathmandu had wealthy "VIP" neighbourhoods, and mapping-listing staff reported substantial scepticism and non-response in these neighbourhoods.

Travel. Mapping-listing staff commuted to clusters via bus, rickshaw, motorbike, and foot. In Kathmandu, most staff never travelled more than one hour to a cluster, however a team working in peri-urban Kathmandu spent three hours commuting one way to one cluster due to the absence of buses or taxis. In

Dhaka, where traffic is notoriously bad, commute times to clusters ranged from 1.5 to 3 hours. Across the three cities, mapping-listing staff recommended hired vehicles to save time.

Area-microcensus versus two-stage clusters. Mappers-listers in Kathmandu reported different experiences in area-microcensus and two-stage clusters. The two-stage clusters were, by definition, ten times the size of area-microcensus clusters resulting in extra days of work and more physical barriers to navigate such as hills and rivers. In addition, the two-stage clusters required more information than area-microcensus clusters, resulting in longer interactions and higher levels of scepticism among residents.

Residents in Kathmandu were generally willing to report number of apartments/dwellings per building, however, they were reluctant to specify the number of households per dwelling and to give household head names. In many two-stage clusters, teams approached a business owner on the ground level who gave number of dwellings on the above floors, but refused to give household-level information, and instead directed the mapping-listing staff to the building owner. One way that mappers-listers addressed this challenge was to approach people at a local grocery store, and start a conversation away from their building. In this context, residents were less likely to feel they were speaking on behalf of the landlord.

Technology. Across sites, mapping-listing staff faced challenges with the tablet applications. While some challenges could have been averted with more, or better, training, other challenges were inherent to the tools and protocols used. First, although OpenStreetMap was updated by mappers-listers before visiting clusters, the updates in various applications occurred on different schedules resulting in different versions of the same map in the field. Specifically, updates to ArcGIS (from which field maps were printed), GeoODK (to collect building GPS points during the listing), and OSMAnd and MAPS.ME (used for navigation) were updated 1 to 30 days after a change was made to OpenStreetMap.

A second problem was the number of unintegrated applications that the mapping-listing staff were expected to use, resulting in lost time and confusion. Despite having multiple navigation applications and a paper map, mappers-listers in all cities reported delays and difficulty navigating to clusters. Once in a cluster, however, mappers-listers did not report challenges identifying cluster boundaries, despite their blocky shapes. Mappers-listers also found recording the listing data in GeoODK was arduous, and they often took notes on paper when speaking to residents and then entered information into the tablet immediately after.

Third, the location precision within OSMAnd and GeoODK were poor, often showing a circle up to 36 metres in which the tablet could be located. Location precision was a particular problem in high density areas (presumably with tall buildings blocking or refracting signals), and resulted in a few instances of a mapping-listing team starting their work, and then realizing that they were recording data one or two streets away from the cluster.

DISCUSSION

By comparing DHS/MICS and SUE household definitions, and area-microcensus and two-stage sampling, we found evidence that standard household survey methods unintentionally omit single adults and non-family households, both of which are more likely to represent disjoined households, or be mobile compared to stable nuclear family households.[13,30,35] This is among the first studies in a LMIC context to evaluate under-coverage due to survey design and methods in face-to-face surveys; such studies tend to be conducted in high-income countries.[14,36]

Although the same protocols and household definitions were used to identify households in Kathmandu's area-microcensus and two-stage arms, the quality of the household listing data appeared better in area-microcensus clusters where interviewers listed households. Interviewers had more skills to interact with the public and substantially more time at each building while administering questionnaires (2.5 to 3 hours per household) to build rapport with residents and learn about atypical and informal housing arrangements. Indicator design effects point to another possible benefit of the area-microcensus design. Although one might expect larger design effects in area-microcensus clusters because near neighbours are assumed to be more similar than far neighbours, the DEFTs for slum, migration, and education indicators in area-microcensus clusters were smaller than in two-stage clusters. This might indicate better coverage of the heterogeneous mix of urban residents, and better identification of atypical and "hidden" households. Other argue that standard household definitions are no longer suitable in complex LMIC cities; rather, individuals and communities are more appropriate units of measurement.[5,35] Further research is need to evaluate potential trade-offs and benefits of moving the household listing responsibility to interviewers using area-microcensus survey designs, but our findings suggest multiple benefits.

Without urban strata, the two-stage sample in Kathmandu was better able to measure tent and shack dwellers than the area-microcensus sample, likely due to the larger area of two-stage clusters. The only way to ensure representative surveys of shack/tent dwellers and other vulnerable populations concentrated in slums is to treat deprived/not-deprived areas as strata, in both area-microcensus and two-stage designs.[16]

We found that response rates in area-microcensus clusters were lower than in two-stage clusters. This may have been due to the greater proportion of vulnerable and mobile households identified in area-microcensus clusters if they were less willing to participate, more likely absent, or felt disempowered to respond. Readers who are interested in area-microcensus survey designs should take account of lower response rates and potentially higher design effects when calculating sample size. The surveys conducted in Dhaka and Hanoi focused on vulnerable and mobile communities, so rates of exclusion may have been higher than the general population.

Societal changes, particularly rapid urbanization in LMICs, have likely caused decay in survey data accuracy due to increased complexity in living arrangements, urban disparity, and population mobility. Not only are vulnerable and mobile populations more likely to be intentionally excluded from surveys, they are at increased risk of unintentional, unmeasured exclusion, and their data are masked in urban averages when they are sampled. Given the importance of household survey data to policy-making, planning, and monitoring progress toward development goals, it is time to evaluate new survey tools and protocols that ensure inclusion of all households.

REFERENCES

1. Bakewell O. 'Keeping them in their place': The ambivalent relationship between development and migration in Africa. *Third World Q* 2008;**29**(7):1341—1358. doi:10.1080/01436590802386492
2. Castles S. Twenty-first-century migration as a challenge to sociology. *J Ethn Migr Stud* 2007;**33**(3):351—371. doi:10.1080/13691830701234491
3. UN-Habitat. Urbanization and Development: Emerging Futures. World Cities Report 2016. Nairobi: UN-Habitat; 2016. http://wcr.unhabitat.org/wp-content/uploads/2017/02/WCR-2016-Full-Report.pdf. Accessed October 24, 2019.
4. UN Department of Economic and Social Affairs (UN-DESA). World urbanization prospects: The 2019 revision. New York, NY: UN-DESA; 2019. https://population.un.org/wup/DataQuery/. Accessed January 13, 2020.
5. Campbell L. Working with People and Communities in Urban Humanitarian Crises. London: ALNAP; 2017. https://www.alnap.org/system/files/content/resource/files/main/alnap-urban-people-and-communities-2017.pdf. Accessed October 24, 2019.
6. Lucci P, Bhatkal T, Khan A. Are we underestimating urban poverty? *World Dev* 2018;**103**:297—310. doi:10.1016/j.worlddev.2017.10.022
7. Vlahov D, Freudenberg N, Proietti F, et al. Urban as a determinant of health. *J Urban Heal* 2007;**84**(Suppl. 1):16-26. doi:10.1007/s11524-007-9169-3
8. UN Statistics Division (UNSD). Household Sample Surveys in Developing and Transition Countries. New York, NY: United Nations; 2005. https://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf. Accessed October 24 2019.
9. ICF International. Demographic and Health Survey Sampling and Household Listing Manual. Calverton, MD: ICF International; 2012. https://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf. Accessed October 24, 2019.
10. Multiple Indicator Cluster Surveys. MICS5 Tools: Manual for Mapping and Household Listing. New York, NY: UNICEF; 2013. http://mics.unicef.org/tools?round=mics5. Accessed October 24, 2019.
11. Grosh ME, Munoz J. A Manual for Planning and Implementing the Living Standards Measurement Study Survey. LSMS Working Paper No. 126. Washington, DC: The World Bank; 1996. http://documents.worldbank.org/curated/en/363321467990016291/A-manual-for-planning-and-implementing-the-living-standards-measurement-study-survey. Accessed October 24, 2019.
12. ICF International. Available datasets. The DHS Program. Rockville, MD: ICF International; 2017. https://dhsprogram.com/data/available-datasets.cfm. Accessed October 24, 2019.
13. Eckman S. Do different listers make the same housing unit frame? Variability in housing unit listing. *J Off Stat* 2013;**29**(2):249—259. doi:10.2478/jos-2013-0021
14. Groves RM, Lyberg L. Total survey error: Past, present, and future. *Public Opin Q* 2010;**74**(5):849—879. doi:10.1093/poq/nfq065
15. Lilford RJ, Oyebode O, Satterthwaite D, et al. Improving the health and welfare of people who live in slums. *Lancet* 2017;**389**:559—570. doi:10.1016/S0140-6736(16)31848-7
16. UN Department of Economic and Social Affairs (UN-DESA). Sustainable Development Goals. Sustain Development Knowledge Platform. New York, NY: UN-DESA; 2018. https://sustainabledevelopment.un.org/sdgs. Accessed October 24, 2019.
17. Research Center for Employment Relations. Living Wage Report: Urban Vietnam. Ho Chi Minh City: ERC; 2016. https://www.isealalliance.org/sites/default/files/resource/2017-12/Urban_Vietnam_Living_Wage_Benchmark_Report.pdf. Accessed October 24, 2019.
18. Elsey H, Poudel AN, Ensor T, et al. Improving household surveys and use of data to address health inequities in three Asian cities: Protocol for the Surveys for Urban Equity (SUE) mixed methods and feasibility study. *BMJ Open* 2018;**8**(11):e024182. doi:10.1136/bmjopen-2018-024182

19. European Commission Joint Research Centre (EC-JRC). Global human settlement city model (GHS-SMOD). Ispra: EC-JRC; 2017. http://ghsl.jrc.ec.europa.eu/faq.php. Accessed January 13, 2020.
20. WorldPop. Population / Individual countries 2000-2020. https://www.worldpop.org/geodata/listing?id=29. Published 2017. Accessed February 2, 2017.
21. Stevens FR, Gaughan AE, Linard C, Tatem AJ. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One* 2015;**10**(2):e0107042. doi:10.1371/journal.pone.0107042
22. Thomson DR, Stevens FR, Ruktanonchai NW, Tatem AJ, Castro MC. GridSample: An R package to generate household survey primary sampling units (PSUs) from gridded population data. *Int J Health Geogr* 2017;**16**:25. doi:10.1186/s12942-017-0098-4
23. Flowminder Foundation. GridSample.org. Southampton: Flowminder Foundation; 2019. www.gridsample.org. Accessed January 13, 2020.
24. Thompson SK. Adaptive cluster sampling. *J Am Stat Assoc* 1990;**85**(412):1050—1059. doi: 10.2307/2289601
25. Thomson DR, Bhattarai R, Dhungel R, et al. Planning team guide v1.7. Surveys for Urban Equity (SUE). Leeds: University of Leeds; 2018. https://medicinehealth.leeds.ac.uk/downloads/download/95/planning_team_guide. Accessed January 13, 2020.
26. OpenStreetMap contributors. OpenStreetMap Base Data. www.openstreetmap.org. Published 2000. Accessed November 12, 2018.
27. ESRI. ArcGIS Release 10. ESRI; 2018.
28. GeoMarvel. GeoODK. http://geoodk.com/. Published 2017. Accessed June 9, 2017.
29. Tumbe C. Missing men, migration and labour markets: evidence. *Indian J Labour Econ* 2016;**58**(2):245—267. doi:10.1007/s41027-016-0017-4
30. Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangaeau R. Survey Methodology. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.; 2009.
31. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013;**13**(117):1—8. doi:10.1186/1471-2288-13-117
32. Ensor T, Bhattarai R, Manandhar S, et al. From rags to riches: Assessing poverty and vulnerability in urban Nepal. *PloS One* 2020;**15**(2):1—19. doi:10.1371/journal.pone.0226646
33. ICF Macro. Training Field Staff for DHS Surveys. Calverton, MD: ICF Macro; 2009.
34. Carr-Hill R. Missing millions and measuring development progress. *World Dev* 2013;**46**:1—15. doi:10.1016/j.worlddev.2012.12.017
35. Oya C. Who counts? Challenges and biases in defining 'households' in research on poverty. *J Dev Eff* 2015;**7**(3):336—345. doi:10.1080/19439342.2015.1068358
36. Eckman S, Kreuter F. Undercoverage rates and undercoverage bias in traditional housing unit listing. *Sociol Methods Res* 2013;**42**(3):264—293. doi:10.1177/0049124113500477

ACKNOWLEDGEMENTS

We would like to thank the city and government authorities in Hanoi, Dhaka and Kathmandu who supported this study in their cities. We would also like to thank Forrest Stevens and Andrea Gaughan who generated a revised WorldPop gridded population dataset for the Nepal sample frame, and Bao Ngoc who contributed to early planning work for the Vietnam survey.

FUNDING

This study was funded by the Medical Research Council (MRC) Global Challenges Research Fund (GCRF) grant ID: MR/P024718/1. Dana R Thomson additionally received funding from the UK Economic and Social Research Council (ESRC) grant ID: ES/5500161/1.

COMPETING INTERESTS

The authors declare no competing interests.

DATA SHARING

De-identified participant data and a data dictionary defining each field is available upon request with ethics approval. Please submit requests to Joseph Paul Hicks (J.P.Hicks@leeds.ac.uk) and Helen Elsey (helen.elsey@york.ac.uk).

AUTHOR CONTRIBUTIONS

DRT, SB, HW, SM, RH, HVM, TE, and HE designed the study. DRT, SM, CC, and HE performed the literature search. Figures were developed by DRT, JPH, and ANP. Data were collected by SK, SM, RB, RD, SG, JF, NJU, TF, and DMD. Data analysis was performed by DRT, RB, JPH, RAS, KQL, and ANP. DRT wrote the first draft, and all co-authors reviewed and approved the final manuscript.