PiMA Survey Design and Methodology

Mudhai, O. F. , Lopes, C. A. , Mitullah, W. , Simutanyi, N. , Fraser, A. , Milapo, N. , Mwangi, S. and Srinivasan, S.

UNIVERSITY OF CAMBRIDGE

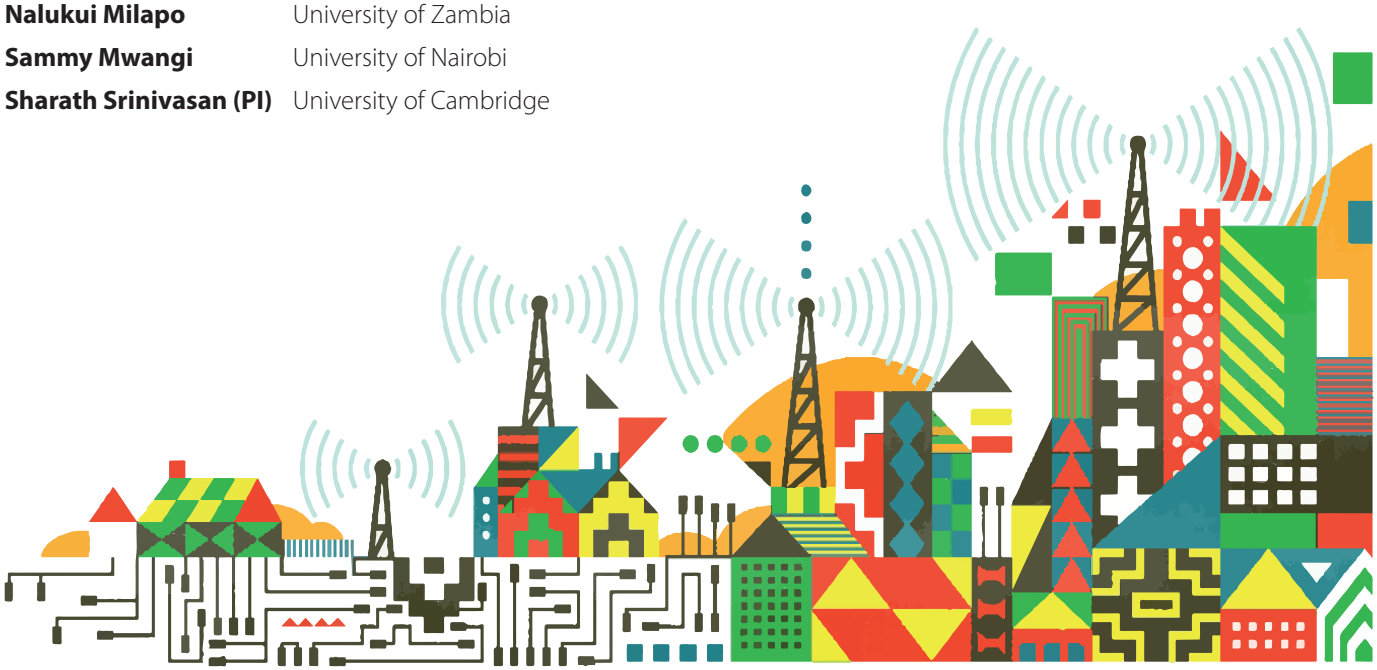
PiMA

Politics and interactive Media in Africa

Working Paper #1

PiMA Survey Design and Methodology

Okoth Fred Mudhai	University of Cambridge
Claudia Abreu Lopes	University of Cambridge
Winnie Mitullah	University of Nairobi
Neo Simutanyi	Centre for Policy Dialogue, Zambia
Alastair Fraser	University of Cambridge
Nalukui Milapo	University of Zambia
Sammy Mwangi	University of Nairobi
Sharath Srinivasan (PI)	University of Cambridge



Project partners

CG+HR
CENTRE OF GOVERNANCE AND HUMAN RIGHTS

University of Nairobi

University of Zambia

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

UKaid from the British people

Research jointly supported by the ESRC and DFID

PiMA Working Papers

The PiMA Working Papers are a series of peer-reviewed working papers that present findings and insights from Centre of Governance and Human Rights' (CGHR) *Politics and Interactive Media in Africa* (PiMA) research project (2012-14).

The project, jointly funded by the ESRC and DFID, focuses on expressions of 'public opinion' in broadcast media via new information and communication technologies (ICT) such as mobile phones in Kenya and Zambia. PiMA examines the political implications of such interactions in the two African countries, with a view to drawing conclusions of wider significance to practitioners and policymakers.

Series Editors: Sharath Srinivasan, Stephanie Diepeveen

Publisher: Centre of Governance and Human Rights, University of Cambridge

Contact:

E: cghr@polis.cam.ac.uk
T: (+44) (0)1223 767 257
W: cghr.polis.cam.ac.uk

Abstract:

The Politics and Interactive Media in Africa (PiMA) project investigates participation in radio and television programmes through new information and communication technologies (ICTs) as it is shaped by the political, social and media context in Zambia and Kenya, as well as the implications of such interactions for democratic governance and poverty alleviation. The empirical research combines different methodologies, including interviews with key informants, focus groups with audience members, in-studio observations and surveys of the general population. This paper describes the methodology for conducting surveys in four sites in Kenya and Zambia. The objective of the surveys was to obtain representative samples of two constituencies per country. Constituencies were selected according to their social and economic characteristics, in order to capture a wide variety of contexts. A random procedure was deployed in all stages of sampling, ensuring representativity of households and individuals of voting age in the four constituencies. The results of the survey can be generalised to the particular constituencies with a margin of error of approximately minus or plus 5% for a 95% confidence interval.

Rationale and Justification

One of the research strategies of the *Politics and Interactive Media in Africa* (PiMA) project was to conduct opinion surveys to supplement other research methods, such as participant observation, key informant interviews and focus-group discussions. The surveys, based on representative samples of citizens of voting age in selected sites in Zambia and Kenya, enabled us to provide indicative answers to the project's first research question: "Which citizens are able to exercise voice using ICTs in media-driven public discussion and opinion-making?" The rationale for conducting the surveys was: (1) to generate demographic information of people who listen to or watch, and text or call-in to radio or television stations; and to understand; (2) the factors that encourage or discourage them from participating; (3) their socio-political attitudes; and (4) their opinions regarding the efficacy of participating in interactive broadcast shows as a way of expressing their political agency or holding authorities to account. The surveys in four constituencies in Kenya and Zambia followed the same core methodology, despite some minor adjustments made in order to take context specificities into account. This paper describes the methodology of the survey, starting by presenting the overall sampling methodology and the sampling procedures in the two countries, followed by descriptions of the strategy for questionnaire development, the procedure for fieldwork training and the process of data cleaning.

General Methodology

Data collection for the PiMA surveys took place during May 2013 (Kenya) and June-July 2013 (Zambia). In Kenya, surveys were conducted in Ruaraka: a peri-urban constituency in the capital city Nairobi, with mixed demographics including one of the city's major slums; and Seme: a rural constituency settled around Lake Victoria in a largely fisher-agricultural community in the western Kenyan city of Kisumu. In Zambia, the surveys were conducted in Mandevu: an urban constituency in the capital city Lusaka with a mixed demographic including some of the city's major slum settlements; and Chipangali: a rural constituency in the country's largely agricultural Eastern Province. The four samples were designed as representative cross-sections of all households in those constituencies. Although no claim is made that the constituencies themselves were representative of the wider national population, they were selected based on the possibility of capturing variation in terms of socio-economic factors, political context and media landscape.

A multi-stage sampling approach was deployed in the four sites, which involved selecting geographically defined units of decreasing size at each stage. The main four stages of the sampling strategy were: (1) cluster sampling for selection of wards; (2) simple random sampling for selection of enumeration areas (EAs) within wards; (3) systematic random sampling for selection of households within EAs ("random walk"); and (4) simple random (Kenya), or stratified by age and gender (Zambia) sampling for selection of individuals within households. Because there were no available lists of voting individuals living in those constituencies based on census data, the population was grouped into units from which reliable data was available, such as EAs. The lists of EAs constituted the sampling frame from which the primary sampling units (PSUs) were randomly selected. In Stages 2 and 3, selection was performed with probabilities proportional to population size. The purpose was to guarantee that more populated areas (wards, EAs) had a proportionally higher probability of being included in the sample. Within each household, individuals were selected using a random procedure. By employing random techniques in all stages of sampling, and using sampling with probability proportional to the population, it may be assumed that all individuals of voting age (18 and over) living in those four constituencies had a known and above zero chance of being included in the sample.

The results of the survey allow inferences to the voting population in the four constituencies (macro-units) with some degree of accuracy (but not to the two countries). The sample sizes are 760 for Kenya (383 for Ruaraka and 377 for Seme) and 688 for Zambia (327 for Mandevu and 361 for Chipangali). The margins of error for a 95% confidence level are no more than plus or minus 5% for both Ruaraka and Seme, 5.41% for Mandevu and 5.12% for Chipangali. The response rate for Kenya was 90.4% (84.6% for Ruaraka and 96.3% for Seme). The response rate for Zambia was not available because the team did not record the number and reasons of unsuccessful calls.

Sampling Methodology

Sampling Design in Kenya

The total sample size for the survey in Kenya was 760 respondents from two counties: Nairobi and Kisumu. The survey took place in one constituency from each county – one peri-urban and one rural. Selection was based on: (1) the proximity to a PiMA case [broadcast] station and urbanity, (2) population political/development profiles, and (3) economic-poverty considerations. The two constituencies, Ruaraka (Nairobi) [1] and Seme (Kisumu)[2], were purposively selected based on their profiles. These two constituencies were the populations from which two samples were drawn. The survey was based on multi-stage sampling comprising four stages:

1. Selection of two or three wards within constituency (purposively clustering sampling);
2. Selection of 48 EAs per constituency proportional of selected wards' population - eight households per EAs (standard practice for Afrobarometer);
3. Selection of household within the wards (systematic random sampling or random walk);
4. Selection of individual within the household (random sampling).

The first sampling stage involved clustering wards within each constituency based on the same three criteria indicated above and their location relative to one another. In Ruaraka, there were three clusters: (1) Baba Dogo and Lucky Summer; (2) Utalii and Mathare North; and (3) Korogocho. In Seme, there were two clusters: (1) West Seme, North Seme and East Seme; and (2) Central Seme. One ward per cluster was purposively selected, taking into account the population size (as heterogeneous as possible) and the number of wards per constituency (cf. Table 1).

Secondly, in each constituency 48 EAs were chosen through the Kenya National Bureau of Statistics (KNBS) with the number of EAs proportional to population size of individuals of voting age within the selected wards. The EAs were the primary sampling units as the smallest well-defined geographic unit for which reliable population data were available. Within each EA, eight households were selected. Selecting EAs ensures more logistic efficiency and reduces the costs of contacting the sample (the travels costs are lower compared to travelling to eight households in different EAs). However, grouping households into EAs decreases the precision of the survey, as people who live in the same geographical cluster tend to be more likely to have similar worldviews and to belong to the same sociological categories.

Thirdly, in each EA, households were randomly selected through systematic random sampling, or random walk. After establishing a starting point,3 the fifth or tenth household, depending on the number or density of households in the EA (urban: every ten and rural: every five), was selected. In Ruaraka, it was mainly the tenth household while in Seme it was mainly the fifth. This involved walking in one designated direction away from the start point, counting houses on both the right and the left (and starting with those on the right if they are opposite each other) and selecting the nth (fifth or tenth) household for the first interview. The same procedure was followed for selecting the subsequent households.

1 Korogocho slum, where Koch FM is located, falls within this constituency. According to data from the Independent Electoral and Boundaries Commission (2014), Nairobi County registered voters for 2013 = 1,732,288 (total number who voted 1,410,663): Ruaraka Constituency = 89,647 registered voters.

2 Seme was until 2012 part of former Kisumu Rural constituency (http://kisumurural.com/), and had typical characteristics of a rural setting. According to IEBC (2014), Kisumu County registered voters for 2013 = 386,606 (total number who voted 348,969): Seme = 35,975 registered voters.

3 A starting point could be a corner, junction, shop, church, school, market, clinic, community centre, headman's house, bus station, etc. At all times considerations were made to ensure that an object for identifying a starting point was not repeated to avoid bias.

Table 1: Demographical profiles and clustering of wards in Ruaraka and Seme constituencies		
Constituency	**Ward Cluster**	**Cluster Characteristics**
Ruaraka (MP – T. J. F. Kajwang') IEBC reg 18+ ppn = 92,261	1. ***Baba Dogo*** (20,000[1]) and ***Lucky Summer*** (13,986)	• Located next to each other – Koch FM catchment. • Mixed accommodation/settlement: – Baba Dogo: Older, industry-based. – Lucky Summer: Newer, expansive, open.
	2. ***Utalii*** (21,000[2]) and Mathare North (25,015)	• Located next to each other – Koch FM catchment. • Mixed accommodation/settlement: – Utalii: Mixed income, institutions-dominated. – Mathare North: Low-income, high-rise private – flats.
	3. ***Korogocho*** IEBC est 18+ ppn = 19,555	• Located at the core of Koch FM catchment. • Slum accommodation/settlement. – Resettlement area so slightly better planned • the residents came from Riverside, River Rd, Grogan.
Seme (MP – Dr J. Nyikal) IEBC est 18+ ppn = 46,063	1. ***West Seme***, North Seme and East Seme IEBC est 18+ ppn: West = 13,266 North = 11,864 East = 10,111	• Located next to each other – Radio Nam Lolwe catchment. – West Seme neighbours North Seme, which neighbours East Seme – closest to Nam Lolwe. • Economy/livelihood – Served by common rural amenities but no water except for West Seme. – Market centres: • North Seme – Bar Korwa • East Seme – Holo [partly in Kisumu West] • West Seme – Smaller centres Akado [shared with Rarieda, Siaya] and Reru.
	2. ***Central Seme*** IEBC est 18+ ppn: 10,822	• Located centrally – Radio Nam Lolwe catchment. – Neighbours all the other three Seme wards. – Market centre: • Kombewa [divisional hq]. • Economy/livelihood – Served by common rural amenities but no water. – Market centre – Kombewa.

Fourthly, in a selected household where consent was gained, a fieldworker listed all members living in the household who were aged 18+ years. From the number of individuals listed, the fieldworker presented cards corresponding with the number of householders listed.[4] A household host (the person who welcomed the fieldworker) picked one card and the household member whose name corresponded to the number chosen was the person to be interviewed. If the person randomly selected was not available after two visits on the same day or refused to participate in the interview, the fieldworker walked away from the household, recorded it on the table of Reasons for Unsuccessful Calls, and substituted the household at an interval of fifth or tenth household.[5] When consent was secured with the selected respondent, the interview proceeded. The duration of the interview was on average (5% trimmed mean) 31 minutes (standard deviation [SD]=6) in Ruaraka and 29 minutes (SD=13) in Seme.

The total response rate was 90.4% (84.6% for Ruaraka and 96.3% for Seme). These figures are based on refusals only. People refusing to answer the questionnaire was the main reason for unsuccessful calls in Ruaraka (73.8% of unsuccessful calls), whereas in Seme the main reason was that there were no adults in the household at the time of the call (74.2% of unsuccessful calls).[6] In Seme, people were more willing to be surveyed, with only 22.6% of unsuccessful calls due to refusals. The reasons for unsuccessful calls are listed in the Table 2.

Table 2: Reasons for unsuccessful calls

4 Cards numbered i.e. from 1 to 15 were presented facing down so that numbers could not be seen by the household host.

5 The Table for Reasons for Unsuccessful Calls, alongside Respondent Selection Table, is part of the questionnaire.

6 The fieldwork in Seme coincided with a funeral that many adults were attending, as well as a market day.

	Ruaraka (n=383)	Seme (n=377)	Total
Refused to answer	59 (73.8%)	14 (22.6%)	73 (51.4%)
Person selected was not at home after at least 2 visits	3 (3.8%)	2 (3.2%)	5 (3.5%)
Household empty for the survey period after at least 2 visits	4 (5.0%)	6 (9.7%)	10 (7.0%)
Not a citizen/Spoke only foreign language	3 (3.8%)	-	3 (2.1%)
Did not speak a survey language	1 (1.3%)	3 (4.8%)	4 (2.8%)
No adults in the household	31 (38.8%)	46 (74.2%)	77 (54.2%)
Total	80 (100%)	62 (100%)	142 (100%)

In order to ensure gender representation, a selection procedure was initially deployed to alternate between a male and a female respondent. Hence, if the first respondent were a man, the second would be a woman.[7] However, this gender-based technique was dropped after the pre-test (before the fieldwork effectively started) in favour of completely random sampling at household level because the gender ratio was not the same in both constituencies. From IEBC data of registered voters used in the 2013 general election,[8] the proportion of men to women in Ruaraka and Seme constituencies was 59:41[9] and 45:55[10] respectively, compared to the national average at 49:51 (Table 3).[11] The gender ratio (M:F) observed in the Ruaraka and Seme samples was respectively 50:50 and 47:53. The z-test for one proportion shows that the percentage of males sampled was significantly lower than the proportion of voting male population in Ruaraka [$z(382)=2.37$, $p<.05$] but does not differ from the figures for the voting male population in Seme [$z(376)=1.57$, $p=.117$]. The gender bias in the Ruaraka sample was corrected using population weights.

From a target sample size of 767 (384 for Ruaraka and 383 for Seme) defined for a 5% margin of error for a 95% confidence level – using a formulae for sample size corrected for the population size (Groves et al., 2009) – the final sample size was a total of 760 respondents (383 for Ruaraka and 377 for Seme). Given these figures, the margins of error ended up at the pre-defined level of 5.0% for both Ruaraka and Seme.

Table 3: Population figures for Ruaraka and Seme Constituencies 18+ Years

	Total		Male		Female		M:F proportion	
	2009 census	2013 IEBC	2009 census	2013 IEBC	2009 census	2013 IEBC	2009 census	2013 IEBC
Ruaraka	116,417	89,427	65,030	52,838	51,387	36,589	56:44	59:41
Seme	45,573	35,735	19,454	16,240	26,119	19,495	43:57	45:55
National	19.46m	14,388,781	9.48m	-	9.99m	-	49:51	49:51

Sources: Compiled from IEBC 2013c, KNBS 2010

Sampling Design in Zambia

The total sample size for the survey in Zambia was 688 respondents distributed in two survey sites: Mandevu (n = 327) and Chipangali (n = 361) constituencies in the Lusaka and Chipata districts of Zambia. These two constituencies were selected based on: (1) their poverty profiles (these sites contain some of the highest concentrations of poor people); (2) the proximity to a PiMA case radio station;[12] and (3) expected political variance between the two sites.[13] The survey was based on a multi-stage sampling

7 This method of ensuring gender representation was also used in Afrobarometer surveys (Afrobarometer Network, 2014).

8 Compiled from IEBC ***Age Gender Matrix per Constituency*** (IEBC, 2013b).

9 Ruaraka registered voters total = 89,427 comprising M 52,838 and F 36,589 (IEBC, 2013c)

10 Seme registered voters total = 35,735 comprising M 16,240 and F 19,495 (IEBC, 2013c).

11 National registered voters total = 14,352,545 comprising F 7,048,846 and M 7,303,699. This figure is for biometric registered voters, excluding 36,236 non-biometric (total 14,388,781) (IEBC, 2013a).

12 Being close to the centre of the capital city, Lusaka, Mandevu is located within the broadcast areas of a dense range of stations, including some at which qualitative research for this project was completed: Radio Phoenix, Muvi-TV, Millenium FM, Radio Yatsani, and ZNBC TV and radio stations. Chipangali is a vast and hilly constituency in Eastern Province on the border with Malawi. Many areas of the constituency receive weak signals for radio (and mobile phones). In other places in the constituency, ZNBC can be heard, as can Breeze FM, Radio Maria and Feel Free FM.

13 Mandevu has been considered an electoral stronghold of the Patriotic Front (PF) since 2006. PF became the ruling party in 2011. Chipangali has remained a stronghold of the recently deposed Movement for Multiparty Democracy.

comprising five stages:

1. Selection of 3 wards within each constituency (random clustering sampling);
2. Selection of 2 Census Statistical Areas (CSAs) per ward (random sampling);
3. Selection of two EAs within the CSAs (random sampling);
4. Selection of a maximum of 15 households within the EAs (systematic random sampling or random walk);
5. Selection of individual within the household through stratified sampling (alternating gender and age).

The first stage involved clustering the wards of the two constituencies into three homogeneous groups based on two criteria: income (with residential areas as proxy for income), and distance from social facilities (e.g. hospital, clinics, markets) and infrastructures (e.g. major roads, electricity). For Mandevu, the clustering was based on residential or income considerations, while, for Chipangali, it was based on the distance from the main facilities. One ward was randomly selected within each cluster for a total of three wards per constituency. The number of cases sampled in each ward was proportional to the voting population of wards (Electoral Commission of Zambia, 2011). Secondly, in each ward two Census Statistical Areas (CSAs) were randomly selected based on the lists of the Central Statistical Office (CSOs) lists. Thirdly, in each selected CSA two EAs were chosen. These were the primary sampling units.

Mandevu had a population of 358,788 and voting population of 180,187 (Central Statistical Office, 2012). The constituency was stratified into three categories based on residential areas (proxy for income): high-cost/high-income (urban), medium-cost/middle-income (urban), low-cost or low-income (peri-urban) residential areas.

Table 4: Characteristics of Mandevu's selected enumeration areas

Residential/Income category	**Ward**	**Enumeration area**	**Sample size/Pop. size**
High-cost/high income	Mulungushi	CSA 01 – Olympia Park North	28 /1,717
		CSA 02 - Olympia Park East	28 /1,497
Medium-cost/middle income	Chaisa	CSA 01 - Chaisa	56 /1,987
		CSA 03 - Kamatete	56/1,222
Low-cost/low-income	Ngwerere	CSA 02 - Muntonyo	92/1,842
		CSA 04 - Kalanga	90/1,226
Total Mandevu			**350/9,491**

Chipangali had a total population of 122,916 and voting population of 52,853. The rural constituency was stratified into three categories based on distance from the main facilities and social amenities. The three categories were: 0-10 km, 10-30 km and over 30 km.

Table 5: Characteristics of Chapangali's selected enumeration areas			
Distance	**Ward**	**Enumeration area**	**Sample size/Pop size**
0-10 km	Nthope	CSA 01 - Mwanjangulu	56 /1,041
		CSA 03 - Chitandika	56/1,631
10-20 km	Chipangali	CSA 02 - Chipangali	60/1,242
		CSA 04 – Chamatundu	48/801
Over 30 km	Rukuzye	CSA 01 - Kamwankuku	70/1,203
		CSA 04 – Mgubudu	60/1,116
Total Chipangali			**350/7,034**

Fourthly, in each EA at least 15 households were randomly selected. The number of households selected in each EA was based on the population size of the ward, in such a way that the probability of selecting a household was proportional to the voting population in each ward. The households were selected using a systematic random sampling following a random walk pattern of selecting the 5th household from the starting point, which was a school, market, clinic, community centre, headman's house or bus station. When there was no one eligible in the household or there was no answer, the household was replaced by the next (6th), which then became the new starting point.

Lastly, one respondent in each household was selected for interview, alternating by gender and age group (18-34 and 35 or over). The gender ratio (M:F) in both constituencies was approximately one (cf. Table 2). In the Mandevu and Chipangali samples, the gender ratio was 45.5:54.5 and 47:52, respectively. The z-test for one proportion shows that the percentage of males sampled did not differ significantly from the proportion of the voting male population both in Mandevu [z(326)=1.63, p=.10] and Chipangali [z(360)=0.76, p=.447].

Table 6: Population of Mandevu and Chipangali constituencies 18+ Years (census 2010)				
	Total	**Male**	**Female**	**M:F proportion**
Mandevu	180,187	90,252	89,935	50:50
Chipangali	52,853	25,734	27,119	49:51
National	5,857,806	2,808,098	3,049,708	48:52

The age distribution in the sample differs slightly from the population figures. The proportion of sampled individuals aged 35 and over was 44.3% for Mandevu and 51.5% for Chipangali. The voting population figures are 37.2% for Mandevu and 49.7% for Chipangali (cf. Table 7). The z-test for one proportion shows that the percentage of sampled individuals aged 18 and over does not differ from the proportion of the population in Chipangali [z(360)=0.68, p=.49] but differs in Mandevu [z(326)=2.66, p=.008]. The age bias in the Mandevu sample was corrected using population weights.

Table 7: Population of Mandevu and Chipangali constituencies by age group (census 2010)					
	Total	**Total 18+**	**18-35**	**35+**	**35+/Total 18+**
Mandevu	358,788	180,187	113,228	66,959	0.372
Chipangali	122,916	52,853	26,595	26,258	0.497
National	13,092,660	6,069,753	4,109,750	1,960,003	0.323

The duration of the interview was on average (5% trimmed mean) 19 minutes (SD=9) in Mandevu and 13 minutes (SD=11) in Chipangali.

Questionnaire Development and Structure

The questionnaire was designed collaboratively by PiMA researchers. All questions were original with the exception of the socio-demographical section, question 22 (contact local officials), 50a (satisfaction with democracy), 50b (accountability and governance), 51 (media and transparency) and 57 (economic hardship), which were borrowed from the Afrobarometer questionnaire (Afrobarometer Network, 2014). The questionnaire was divided into three parts: introduction/sampling, core questionnaire and contextual questions.

The first part of the questionnaire included the introduction below, which was read aloud (roughly as a guide) to the host (the person who opened the door) and to the respondent in order to obtain informed consent:

Hello, my name is _____ from _____ University/Institution. We are conducting research today on issues affecting people in Zambia and Kenya. We are writing an academic book about how people use the media and communications technology. I am going door-to-door in this neighbourhood asking people to take part in a survey and would like to ask someone from this household a few questions. The survey will take about 30 minutes to one hour. All of your answers will be kept strictly confidential. Would you be happy to participate?

The first part of the questionnaire also contained the field record for selection of respondents and households, including a table to be completed with geographical details (country, county, constituency, ward, sub-location, village and EA number); a table to list all members of household for random selection; and a table with reasons for unsuccessful contacts. These tables were to be completed by the fieldworkers.

The second part, the core questionnaire, contained 52 questions and was divided into six sections:

1. Mobile phone and internet use;
2. Radio listenership and TV viewership;
3. Contacts to local officials to solve local problems;
4. Participation in interactive programmes and perceptions of their impact;
5. Perceptions of democracy and political participation;
6. Socio-demographical questions.

The third part covered contextual information on the setting and context of the interview, such as the language used in the survey, the respondent's behaviour during the survey, the mother tongue and gender of the fieldworker, and the duration of the interview (Q53-Q58).

The questionnaire was peer-reviewed by five academics in the fields of media-communication, African politics and public opinion, and their feedback was incorporated in the final stage of the questionnaire development. The last version of the questionnaire was "indigenised" by adjusting the wording of questions, the country context (changing country references and country-specific categories such as education levels), and by adding country or constituency-specific questions (listenership of *Africa's Voices* programme in Radio Nam Lolwe in Seme[14]).

The questionnaire was then translated into local languages (Swahili and Luo on paper in Kenya, Nyanja and Bemba in the field in Zambia) by the local teams. The process of translation was guided by the understanding that rather than seeking literal translations; the questions had conceptual equivalence (the meaning of the questions would be equivalent in all languages), as the comparability of results among different languages is only possible when the questions trigger the same cognitive processes and relate to the same meanings. All the questions included the option "Don't Know/Refuse to answer".

Fieldwork Training

Kenya

The survey deployed 12 fieldworkers, in three groups of four, with each group covering 16 EAs and supervised by a senior member of the team (supervisor). The fieldworkers received training for either three or four days, which focused on familiarising them with the questionnaire, sharpening communication and interpersonal skills, and delivering instructions for the selection of households and respondents. The research ethics were discussed, and fieldworkers received training on how to ensure that interviews were confidential and how to obtain the respondents' informed consent. The fieldworkers were given a copy of a line ministry letter authorising PiMA research through Kenya-based Co-investigator Professor Winnie Mitullah. By the end of the training, the fieldworkers had a sense of the purpose of the survey, and their role in selecting the sample and implementing the questionnaire. Each fieldworker completed at least two practice interviews with "real" subjects before leaving for the field soon afterwards. Each

14 CGHR's Africa's Voices project piloted a methodology for public opinion gathering in Africa using interactive radio and mobile phones. An outline of the project is available from the CGHR website: www.cghr.polis.cam.ac.uk/research_themes/pdtm/africas_voices.

supervisor did at least one interview in order to have a better understanding of the challenges in the field. A total of 27 practice interviews were completed in Dagoretti, a peri-urban constituency to the far western side of Nairobi, with both urban and rural characteristics (the test was carried out in the parts of the constituency that were most clearly either rural or urban). These practice interviews also served to pre-test the questionnaire, and as a result some adjustments – mainly in the wording of questions – were made before the fieldwork started. A debriefing session was carried out after the pre-test to ensure that all procedures were well understood before the fieldwork effectively started.

During the fieldwork, the team met at the end of each day to share experiences and align procedures in the administration of questionnaires. However, there were no further substantive changes in the questionnaire. The quality of data collection was monitored through random back-checks of two interviews in each EA; there were no anomalies found in the administration of questionnaires.

Zambia

In Zambia, the survey work involved six fieldworkers who worked in pairs in EAs. Prior to fieldwork, the Zambian core research team trained the research assistants. The focus of the training was to refine interview skills and build understanding of the study objectives and the sampling frame.

A one-day pre-test was conducted in Munali constituency (M'tendere and Kalingalinga ward), close to Mandevu, in which 36 questionnaires were administered. Supervisors reviewed the performance of the fieldworkers and discussed observations from the pre-test with the fieldworkers.

The fieldwork took a total of 10 days in each constituency (Mandevu and Chipangali). During this time, the interviewers met and the completed questionnaires were checked on a daily basis. Due to time constraints, it was decided that random back-checks would be done at the end of the survey.

Data entering and Cleaning

The answers to the questionnaire were entered into SPSS datasets by the Kenya and Zambia teams using the same template. The variables were correctly labelled and missing answers were given the value of -1 (true missing) or 88 (not applicable). The hard copies were numbered from 1 to n in both countries and these numbers entered into the dataset as the ID variable.

All the open-ended and the "other" answers were entered verbatim into the dataset. The open-ended answers were coded using a common codebook created by the team in Kenya and passed onto the team in Zambia. The codes cover at least 90% of all possible answers for open-ended questions, i.e., the answers were coded in such a way that the frequency of the "other" category was kept at a maximum of 10% of the total answers to that question. The "other" answers in closed-format were coded using the same rule, meaning that if the frequency of the category "other" was at least 10% of the answers to that question, new categories were created and included in the set of categories for that question.

The data entry process was validated by checking 25% of the questionnaires against the values in the database, and mismatches were corrected. One questionnaire presented mismatches in Zambia. No mismatches were found in Kenya. In Kenya, one supervisor managed six data entry clerks, and in Zambia, one supervisor managed three data entry clerks.

The dataset was cleaned by the local teams by running frequencies for each variable to check for out-of-range or system missing data (basic data cleaning), and also running cross-tabulations to check for internal consistency of linked questions (advanced data cleaning). The data was re-entered by the local teams when errors were found. The PiMA team in Cambridge performed an independent assessment of the quality of the cleaning before analysis of the datasets began.

References

Afrobarometer Network (2014), *Round 6 Survey Manual*. Retrieved from www.afrobarometer.org/files/documents/survey_manuals/survey_manual_r6_english.pdf.

Central Statistical Office (2012), *2010 Census of Population and Housing: Population Summary Report*. Retrieved from www.zamstats.gov.zm/report/Census/2010/2010%20Census%20of%20Population%20Summary%20Report%202-%20Final.pdf. Lusaka: Central Statistical Office.

Electoral Commission of Zambia (2011), *Registered Voters per Polling District*.

Groves, R. N., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2009), *Survey Methodology*. Hoboken, NJ: John Wiley and Sons.

Independent Electoral Boundaries Commission (2013a), *Summary of all Elective Positions in the March 4th 2013 General Election.* Last update 18 July 2013. Retrieved from www.iebc.or.ke/index.php/resources/downloads/item/summary-of-all-elective-positions-in-the-march-4th-2013-general-election?category_id=32. Nairobi: Independent Electoral and Boundaries Commission.

Independent Electoral Boundaries Commission (2013b), *Age Gender Matrix per Constituencies*. Last update 27 February 2013. Retrieved from www.iebc.or.ke/index.php/media-center/press-releases/item/age-gender-matrix-per-constituency. Nairobi: Independent Electoral and Boundaries Commission.

Independent Electoral Boundaries Commission (2013c), *Age and Gender Matrix per Constituency.* Last update 2 February 2013. Retrieved from IEBC in February 2013 [no longer available online at the time of publication]. Nairobi: Independent Electoral and Boundaries Commission.

Independent Electoral Boundaries Commission (2014), *4th March General Election: Election Data*. Last update 5 June 2014. Retrieved from www.iebc.or.ke/index.php/resources/downloads/category/publications. Nairobi: Independent Electoral and Boundaries Commission.

Kenya National Bureau of Statistics (2010), *Population and Housing Census 2009: Volume 1C – Population Distribution by Age, Sex and Administrative Units*. Retrieved from www.knbs.or.ke/index.php?option=com_phocadownload&view=category&id=109:population-and-housing-census-2009&Itemid=599. Nairobi: Kenya National Bureau of Statistics.

PiMA Working Papers

The **Centre of Governance and Human Rights (CGHR)**, launched in late 2009, draws together experts, practitioners and policymakers from the University of Cambridge and far beyond to think critically and innovatively about pressing governance and human rights issues throughout the world, with a special focus on Africa. The Centre aims to be a world-class interdisciplinary hub for fresh thinking, collaborative research and improving practice.

Series Editors: Sharath Srinivasan, Stephanie Diepeveen

Publisher: Centre of Governance and Human Rights, University of Cambridge

Contact: cghr@polis.cam.ac.uk, (+44) (0)1223 767 257

For further information please see the CGHR website:
www.cghr.polis.cam.ac.uk/research_themes/pdtm/pima

Citation: Mudhai, O. F., Abreu Lopes, C. et al (August 2014). 'PiMA Survey Design and Methodology', PiMA Working Paper #1, Cambridge: University of Cambridge, Centre of Governance and Human Rights. ESRC/DFID Award #ES/J018945/1.

Centre of Governance and Human Rights
POLIS · 7 West Road · Cambridge · CB3 9DT · United Kingdom
www.polis.cam.ac.uk/cghr