# Collecting social data: Adapting to the COVID-19 pandemic

*Agenda*

Tuesday, 13 August 2020
10:00-12:00

**Mariëtte Croukamp, Ask Afrika**

*Methods and madness during COVID-19*

**Shirley Jeoffreys-Leach, Petronella Tizora and Bobby Berkowitz, FinMark Trust**

*Implementing and interpreting multi-country longitudinal studies through the COVID-19 pandemic*

**Kim Ingle and Timothy Brophy, SALDRU, University of Cape Town**

*Experience from re-engaging with an existing face-to-face panel via CATI in a rapid response survey*

*Reflection:*

*What are we compromising and what are we gaining?*

# Mariëtte Croukamp, Ask Afrika – Q&A

1) Julia de Kadt: I'd love to hear a bit about how you set up your panel & your CATI sample and data frame. Frame obtained from a secondary source; details of 30 millions South Africans. Randomly draw on a weekly basis to ensure the same person isn't contacted repeatedly. The panel is an existing panel, that covers different demographics & areas. This gives another way to reach people, and assists with response rates.

2) Tara Polzer Ngwato: Please say more about what you mean by 'fractals'. This is another term for what are effectively themes – for example corruption, school openings, mask-wearing behaviours, behavior change and adaptation in response to information from different sources.

3) Julia de Kadt: I'm really interested to learn about how your team is coping with running a project over such a long and difficult period - how has the team been managing burnout and stress? Has become easier over time, as team 'fitness' has improved. Have a routine of daily tasks, with strong project management in place. Quite stressful and hectic in beginning, but with good rhythm and clear expectation has become easier. On week 20 now, so a bit of a marathon, but results are empowering, and seeing the data being used is motivating.

4) Pascal Richardson: How will the interview data be used? 9000 interviews is a lot! Will each be transcribed, translated and analysed? (similar question for the voicenotes). The 9000 relates to the quantitative responses so don't need translation/transcription. The responses for the qualitative are all analysed. Using the daily themes, it was fairly easy to extract data from the 10 respondents. Some of this was incorporated into a video, and it was also used to substantiate findings from the quantitative work. The data is used by Ask Afrika, and results are available on the online portal. Hoping to see widespread use of results to make a positive difference in the lives of South Africans.

5) Tara Polzer Ngwato: do panel respondents receive incentive payment? Yes, as with any panel, but the incentive is quite small.

6) Ben Roberts: Is the data going to be made available for secondary analysis via the portal? The data isn't on the portal, but has been made available to some analysts. Anyone interested can contact Ask Afrika.

7) Jeremy Seekings: Sampling: What is your response rate on CATI? Don't have an exact figure to hand, but definitely above 80%. Very good response rate because people are very keen to share their experiences. The sampling frame is working quite well.

8) Jeremy Seekings: Sampling: do you have any empirical data on the weaknesses of your realized sample? Quotas and weighting assume that responses within undersampled groups were random, which is implausible. Did find some over/under sampling in sub-groups, and correct for that week-on-week. Because the full research cycle takes place in a single week, have to accept some under/oversampling. For example, with Mpumalanga an undersample of 1% and 2% for North West. More information is available on request.

# Shirley Jeoffreys-Leach and Bobby Berkowitz, FinMark Trust – Q&A

1) Julia de Kadt: I'd be interested in your thoughts about whether/how the reach of these various modes shifts as questionnaire length increases? Obviously face to face allows for a longer survey. With SMS particularly, when people get bored just put phone away. For mixed mode, would want to keep survey the same across modes, or at least keep the first part of the face to face identical with the SMS, so that it's easier to bring data together.

2) Julia de Kadt: And also whether you see any significant variations in reach of subgroups across countries/contexts, and indeed within countries The design effect example summarises this quite well. Definitely across countries and waves. Compare the weighted and unweighted data for questions of interest broken down by subgroups to see how it works out.

3) Tara Polzer Ngwato: Quotas work when there is relatively recent and known representative demographic data for the pop of interest and also some baseline data on the topic of interest from a random sample to compare/weight against. As we move further away in time from representative sample frames and/or look at new questions without points of comparison, what do we do to check our sample validity/bias? Very true – more recent target data is better. Debated which surveys to use extensively upfront. There were some FinScope surveys that were more recent, but finally went with DHS across all countries. A split sample approach meant 50-80% of samples were collected using random digit dialing to get a good spread of people, and then quotas were applied thereafter to make sure under-represented people were adequately covered. There are difficulties with target data – even seeing delays now from StatsSA, but much more extreme in countries like DRC which hasn't had a census since the 90s, and looks very different now. This leaves researchers 'shooting in the dark'. This is very important to note especially with non-representative surveys. This is an important area for work with statistical agencies.

4) Ingrid Reinten: in all the survey's that are done – how many can be used as at subnational level - as a LG cllrs my interest is on knowing a lot of info about my specefic geographic area. For the reasons explained in Q3, the FinMark data can't be used at more local levels. The other surveys presented today are also not really designed for use at more local levels.

5) Julia de Kadt: And with reference to Ingrid's question - would love to hear FinMark thoughts on whether/how the sampling & weighting approach would work at the subnational level. The same applies here – you need appropriate target data, which requires a large enough reference sample. This poses challenges for subnational surveys – you really need something like census. Though for Gauteng the Quality of Life survey is also useful.

6) Tara Polzer Ngwato: How does Rake weighting interact with power calculations in smallish/medium samples (e.g. around 3000 nationally or under) This increases the complexity – the more things you control for, the more complicated your power calculations become.

# Kim Ingle and Timothy Brophy, SALDRU, University of Cape Town – Q&A

1) Julia de Kadt: Could you say a bit more about how the alpha & beta data release process works? And particularly how and where weighting fits into this? Given the NIDS-CRAM timeframes, there were two very internal releases. The alpha release provided the internal CRAM team with a preview of the data structure and the kinds of results coming in. The beta release went to a broader group including all paper authors, and similarly gave a little more time to prepare analysis.

2) Is it possible to break down the response rate a little? In particular what proportion of non-response relates to not being able to reach the right person vs refusals to participate? Possible reasons for NIDS CRAM response rate of 40% compared to AskAfrika's 80% for cold-calling? Different definitions of 'response rate'? The NIDS-CRAM response rate refers to the proportion of sampled individuals who were contacted. Because of very limited time, had to close batches of people off as uncontactable fairly quickly, so that the sampling team could review and generate the next sample group. The vast majority of the 60% non-response was people who couldn't be contacted at all. The last survey was in 2017, so numbers were old, and sometimes didn't have an individual's personal number. Unlike a field-based survey, weren't able to go and look for people. For more information on the responses rates, please look at the technical paper on sampling.

# TO ALL– Q&A

## What from working in this period, methodologically & practically, they'd take forward into future work

**Mariette (Ask Afrika) –** Will definitely retain agile workflow. Have learned that it is possible to do a survey in a really quick turnaround time, while retaining necessary validity and reliability. The mixed methods approach works really well, particularly when you struggle in particular areas. An alternative approach with a CATI component as part of a large-scale household survey is worth considering.

**Shirley (FinMark Trust) –** Creating deeper structure to the work that has happened. Some of the work was very ad hoc, and in hindsight, adding more structure up front would have been worthwhile, even though it would have delayed field entry a bit.

**Bobby (FinMark Trust) –** The purpose of doing wave-by-wave indicator reporting. The diagnostics in ensuring samples are as close to reference populations as possible reduces bias in indicators. This means you can track what's actually going on, and reduce noise. This kind of sample management and the design effect is very useful for tracking.

**Tim (NIDS-CRAM) –** Collaborative nature of work has been very important. A whole group of people came together, and someone was always ready to help. This kind of collaboration is work maintaining.

**Kim (NIDS-CRAM) –** This era of working remotely has shown that remote collaboration tools can be very effectively. Encourage people who are considering this kind of work to make use of a pre-test. Even if you're really tight on time, a pilot or pre-test is important. Especially if you're venturing into new modes and methodologies.