On the record

Randomised household surveys: challenges and considerations

Kelley Moult

kelley.moult@uct.ac.za

As anyone with empirical fieldwork experience knows, even best laid data collection plans rarely go off without a hitch. There is often rich learning from these challenges, although we seldom reflect on them in the literature. This interview asks the University of Cape Town's (UCT) Safety and Violence Initiative (SaVI) director, Guy Lamb, and study coordinator Ncedo Ntsasa Mngqibisa about their experiences carrying out the Gugulethu component of a randomised household survey project that took place in Gugulethu and Manenberg in Cape Town in 2017. Young people between the ages of 12 and 18 years old, and one of their caregivers, were interviewed using a detailed (structured) questionnaire. This project was a partnership between SaVI, Amandla EduFootball and Dr Ian Edelstein from the Human Sciences Research Council (HSRC), and focused on youth resilience, deviance and development. The project was funded by the Department of Cultural Affairs and Sport in the Western Cape Provincial Government.

Kelley Moult (KM): Can you give us some background to the project and what it aimed to do?

Guy Lamb (GL): The project aimed to do a longitudinal survey – following a population in Gugulethu and Manenberg for 10 years – focusing in on a very specific area with a radius of 1 000 km^2. Because we were interested in the dynamics of violence and violence prevention, we wanted to interview a representative sample of youth (both men and women) from that area, but weighting the sample towards young men because violence particularly affects men both as perpetrators and as victims. We were modelling a project by Ian Edelstein from the HSRC that has been successfully carried out in Khayelitsha and so we used a similar method, up-scaling it in Gugulethu and Manenberg.

KM: So you set off with your project model, and almost immediately hit implementation challenges. Can you tell me a little about what happened?

GL: After piloting the questionnaire in the area about a year before, we uncovered some problems with the data that had been collected by the data collection service provider that we had contracted to do the work, including possible fieldworker fraud. The service provider eventually pulled out of the contract, which led us to establish our own team for the Gugulethu data collection, while we engaged another service provider to cover Manenberg. We pulled together a group of experienced field researchers – people with survey experience – to go into Gugulethu to conduct the research. After our fieldworker fraud experience we decided not to

go with our original plan of using a paper-based questionnaire, but instead we acquired tablets and software developed through the HSRC to take into the field.

We used a very strict sampling methodology, which was linked to GPS coordinates where households were preselected in both informal and formal housing neighbourhoods. Households in the formal area were preselected using addresses, and GPS points were selected in the informal settlements as a starting point for the fieldworkers. We had anticipated that this might be tricky, and so we had a very specific strategy that the surveyor needed to follow if they came to a household or to a GPS co-ordinate point, and found that there was no house there, or the household didn't have any family members that conformed to our population sample.

Ncedo Ntsasa Mngqibisa (NM): What this translates into in practical terms is that if you go to a household and you do not find somebody who is between the ages of 13 and 18, you are given another household to visit as a substitute. This is easy among the formal houses, but in neighbourhoods of informal houses, it is much more of a challenge. The substitution plan was that the fieldworker should turn right and go three houses from that household where you didn't find someone to interview. But in some of these neighbourhoods 'three houses' may actually be three shacks which can belong to the same family, which doesn't strictly work in terms of the sampling strategy. So these are some of the things that ordinarily you don't think about when you are planning research in these spaces. You just assume that each shack belongs to each household, but sometimes it doesn't work out like that.

KM: How easy was it to negotiate access to communities and survey participants?

NM: Before you get into these communities you need to negotiate access, and that is not always an easy thing to do. The community in

question may have been over-researched, in part because of its proximity to universities. In our case, there are at least two universities very close by the community we had selected. And so having some background knowledge about their access to universities becomes relevant as you start asking questions in this community. If the people there come from an area that has little or no exposure to university research their understanding of what research is may differ, and so you may need to take longer explaining to them what research *is*, even before you get on to explaining what the research is *about*. If someone has taken part in research before, they already know what you are talking about. So, on the one hand you may run the risk of making them feel like you think very little of them by trying to explain too much or, on the other hand, you may not do enough to secure their participation.

Also, these communities, rightfully so in my view, want to see benefits from all of this research. Interested parties within the community will confront you from the start about how they are going to benefit from the project, and it's a very difficult question because 'building the literature' may not be something that is construed as a benefit to them. One of the ways that this problem can be overcome is to make sure that the research that you do in the community is going to be translated in a way that they are able to use it to further whatever they may demand from government, and which may relate to the subject of the research. The community also asked us where we got our fieldworkers, because one of the major demands from the community was that we needed to ensure that local people got jobs with the project.

It is also very difficult to know who specifically to negotiate access with, who the gatekeeper for the community is. You may meet with a ward councillor and the councillor is not concerned about the project and gives you an easy go-ahead. But as you are meandering about you

meet street committees, or you meet another body that exists in that community, who know nothing about your project, and also need to be consulted. It is therefore critically important to go into the community long before you are actually going to be doing your study. You need a lot of skill to negotiate this access and make them understand that you are only able to commit in the short term.

A challenge here, though, is that projects already come with timeframes that determine when you need to start and finish a project. But we set these timelines without doing a background check to understand who is actually going to be a stumbling block to the project, who to negotiate access with. If I must meet all of these bodies, who do I meet first? What is the protocol here? Sometimes you may even think that because you have met a street committee in one neighbourhood or block, and you are still working in the same geographical area, that the people you have access permission from represent the entire area. But then you come to another street and they say 'you didn't speak to us, you only spoke to those people'. Now you must stop, and you must renegotiate. But now you are negotiating with people who already view you with suspicion simply because they feel that you have ignored them. But you actually didn't – you just did not know that in a small area like that you would have so many interested groups, just to access the community.

KM: Your fieldwork team had a number of other challenges that really make the 'pitfalls' of survey research that are in the literature come to life, for example around language and non-response. Can you tell us a little about those?

NM: These kinds of challenges are not always easily anticipated, and we had to sometimes think on our feet, while at other times we would have to come back to the office and rethink our strategy, and regroup. To face the myriad of challenges we had to retrain our fieldworkers a couple of times.

We thought of the survey as a comparison between two communities: Gugulethu being one community, and Manenberg the other. But we were actually dealing with three communities, because in Gugulethu there are at least two communities that differ very much ideologically, even though they are just a street apart from each other. They even see themselves very differently, split very much along the lines of those who come from the Eastern Cape, and those who are born in the Western Cape. Their language, the isiXhosa that they speak, is different and so they don't all understand your questions, even though you are speaking isiXhosa yourself. So there are problems with translation because you have made an incorrect assumption that because all these people come from the same area they all speak isiXhosa, and so everyone is going to understand your questionnaire.

You cannot assume that because something works in Khayelitsha it should work in Gugulethu. We were perhaps a little bit too relaxed in terms of our preparation because we thought that since the project has worked in one informal settlement it must work in the other informal settlement. Actually these are two different informal settlements. There are people in the neighbourhood who have been chased away from other areas, they do not want to belong here, and so they are not as organised as others. Their mindset about being researched may differ from people in rural areas who may have not experienced research at all. There are small details that can change every day within a particular context: informality within that space, people's senses of belonging and who actually resides here, whether the neighbourhood is made up of people who come from the Eastern Cape. We assume that people are the same everywhere because this is an informal settlement.

We had a lot of challenges enrolling households. In the formal housing community, you don't find anyone at home during the day, because they work and they don't send their children to school in the same community, which poses one kind of access challenge. In the informal community you do find people at home during the day, but these people are also very mobile, which means that today you might find this household member, but tomorrow you cannot find them, they've moved, or their houses have burned down. People also get home late [from work], and their children also only come back around 7 o'clock [in the evening], and then everyone is cooking or preparing for tomorrow. It is not easy for a researcher to come into their home and say 'I'm asking to do research.' People would say 'come back another day', but another day turned into yet another day. From the project's point of view, the plan was only to collect data in the area for a specific amount of time, and going back to households so many times would waste a lot of resources.

KM: It must have been an enormous challenge managing this in the field.

GL: Ncedo was at the coalface of managing the fieldworkers and troubleshooting day to day. But we were constrained by the fact that there was a parallel study happening in Manenberg, and we therefore couldn't stray too far from the sampling strategy that had been devised. We had to negotiate with partners, and sometimes there was a lack of clear understanding or appreciation for the context in which we were operating.

From a management point of view, we had a set budget, which also determined what was possible. Based on models of other survey research projects, we estimated that a field worker would be able to do, on average, three questionnaires a day. Our fieldworkers started feeling guilty and panicking because

they were going to households and not finding respondents, and so they were worried that they wouldn't be paid. We really struggled with what to do. On the one hand, you have a finite amount of money, and you want to get a certain number of completed questionnaires per day, and you want it to be done properly. But the challenges we encountered in the field meant that we didn't hit our targets because it was practically not possible to do so. It wasn't fair to exploit the field workers, and we recognised they needed to be paid fairly, and so we made adjustments in how we paid them. But it was a particularly tricky issue for our project.

KM: You used cell phones as a key part of your strategy – getting people's cell phone numbers as a way to maintaining contact with them. But in South Africa that's a lot more complicated because of cell phone churn. Even though we have among the highest levels of cell phone coverage – in other words, most people have phones – we also know that people don't keep the numbers for very long, or frequently switch between numbers.

NM: Because this is a longitudinal study, we needed to be able to contact the same participants over time. In the formal housing neighbourhoods, you can return to a house, and if that structure looks the same, you can be almost certain that these are the same people that you interviewed three years ago, or if they have moved, that the person who lives here now may actually tell you where they have moved to. But in informal settlements this may not be possible – you may be talking to someone who does not know where these people moved to. That is why it is dangerous for researchers to just copy these things from other countries and paste them into our context. Because our context differs a lot.

Our plan was to enrol participants, and then use cell phone numbers to keep in contact. We realised quickly that particularly among young people, expecting that a 13-year-old would

have the same cell phone number six months down the line is unrealistic. It is very easy to get access to new phone numbers – it costs R1 to get a SIM card – and the cell phone companies offer deals to induce people to use their networks. From our project's point of view that is a problem, because some of our potential participants come from a community where there is a lot of mobility, where people move in and out of a particular space. We hoped to use cell phones as something that we could use to track down our participants over time, but we quickly realised this was going to be impossible because of these challenges that are peculiar to South Africa.

GL: We assumed that by accumulating as many cell phone numbers as possible of the people who are related to the respondent – family members, grandparents and the like – we would be able to stay in touch in one way or another. We didn't anticipate, for example (even after our pilot study), that people would give us fake cell phone numbers because they were concerned about sharing sensitive information with us. Some of our questions were about whether the respondents had committed acts of violence, acts of crime, or done things that they are not supposed to, and people were concerned about the consequences of reporting to us. There was therefore no incentive (from their perspective) to share their numbers with us.

We decided to address this by providing participants with a cell phone voucher for airtime, both as an inducement to participate, but also as a way to verify the cell phone number. This worked in the short term because we could verify the number in that way. But in a longitudinal study, by the time you want to phone the person back for a follow-up interview two years later they may have changed numbers five, six, seven, eight times.

KM: You purchased tablets and software to assist with the data collection. How did this work in the field?

NM: Although we had a number of tech challenges in using the tablets, the technology was helpful in some critical ways. When you went to a household, you pulled up your [survey] tab, and you could report immediately [on what you encountered], for example, you could let the project staff know that you were at a particular household and there was no-one home. Fieldworkers could send the reports right as they happened, which was a real benefit of using technology. To complement this, we also used debrief sessions. Every day we would meet and talk about the day, so that if something happened – for example, a participant had difficulty in an interview – we were able to log it so that when we analysed the data, that information was tagged on to that interview record to give background information to the analyst.

But of course, technology only takes you so far. We didn't always have good data coverage, and even though the tablet might tell us where we are on the map, in terms of safety there are a lot of things happening around us that we do not actually know. You have to balance these very carefully so that the study's data is not compromised, while at the same time you try to ensure that everyone you meet understands the project, and is supportive. You need someone who can tell you 'we don't go down that street', and somebody who is going to know where your fieldworkers are. Luckily, we were able to negotiate such that we even identified (and hired) people who walked around with us to help protect us. But this meant that the science of the study had to be flexible. When you are told by someone that you can't go down a particular street, you need to listen. Even if your sampling plan says you should be going down that street, in reality you can't do so when you are told not to. You don't even try to verify that information because

if somebody says that, it is enough. You don't want to risk people's lives and then end up feeling guilty about it because you were warned and you chose not to heed the warning. But science wants to stick with the sampling rules because you want to cover a particular radius. Sometimes we were told we could not go into whole sections of Gugulethu, and we had to be OK with me calling the team and saying 'we are not going to be covering area B, C and A'. And then we just had to move on to areas where we were able to go. That is how tricky the safety situation was out in the field.

KM: You raise an important issue of safety: safety for researchers and safety for participants in these spaces. Ncedo, when I've heard you talking about this project before you've talked about how your role as project manager changed as you tried to address questions of safety in executing the data collection.

NM: Generally in South Africa, we know that we have a problem with crime. Nyanga is said to be the so-called 'crime capital' and here we were working in Gugulethu, which is within walking distance from Nyanga. In fact, at times you might be thinking that you are in Nyanga, but you are in Gugulethu. So this is the context we were working in. You might also find that while we were doing research in July it was still safe enough to go into a particular area, but by May of the following year it might be totally different. This means that you need to have a cosy relationship with the community, which requires nurturing because your safety depends on it. The community members must actually feel that they want to protect you, that they want to give you the information about the environment that will protect you. You therefore do not compromise on that relationship because to do so means compromising on the safety needs of the study.

In addition, you have to be very mindful of issues like confidentiality because participants'

disclosures of violence perpetration and victimisation can compromise their safety. So this gives you other challenges – for example, when you are talking to a child, the parent must not be present to protect confidentiality. When we talk to the parent, the child must not be present. This may be possible in other spaces, but where the home is just a one-room shack, it is very difficult to say to somebody 'leave your house because we want to conduct this research', especially when the parties may not even see the research as beneficial to them. And so you have to think about the other options that you might have for interview spaces. We thought of using our research vehicle, but now you are seen taking a 17-year-old into the car when you are a male who's 30-something. That doesn't look right in the community and may distract people from your purpose. The community thinks 'we thought these guys were coming here for research, look now what he's doing to young girls'. We are also asking about hard questions – for example, asking about violence and drugs – which make it very difficult to negotiate safety in that environment.

So we decided to work in groups to address safety concerns. If we had to sample two households in Street A, instead of being scattered around, trying to chase participants, we worked in one area. The person who is inside the house knows that someone will be there to pick them up immediately when they come out after the interview. Because otherwise they stand at the gate waiting for their ride with these tablets and cell phones, which makes them targets for being robbed, particularly in winter where nobody is actually outside. So safety was a challenge for us. Instead of being a researcher, I found myself also being a security person who had to drive to each and every house so that when a fieldworker reported that they were finished, I was there with the car as soon as they came out of the household.

GL: We set up safety protocols from the outset of the project, and we had arranged, for example, that the fieldworkers would be wearing clothing that

made them identifiable as part of the project. We had high visibility vests, an identity card, and a letter to show to people whom they were approaching as potential interviewees to vouch for the project. We also spoke to all of the various security stakeholders and gatekeepers within the area. But we still had to really adapt on the fly, developing strategies based on what we encountered. The Gugulethu team was very successful in doing this, whereas the service provider in Manenberg had major difficulties in terms of security and access. This is a community that is just on the other side of the railway line, but it had completely different community dynamics, very high levels of violence, gang violence and suspicion.

NM: Our safety was in many ways related to the project that we were evaluating. Where the project didn't always have a good relationship with the community, and people didn't care about it, the fieldworkers' safety was compromised because the community was less invested in the study's outcomes. This had nothing to do with anything that we could control as fieldworkers. So yes, it does help to have things that will make you identifiable as part of the project, but only when the community actually wants to identify and protect you.

KM: These kinds of challenges are (at least in some measure) acknowledged in the literature on research methods, and in this case, you were working from a model that has previously worked in another community. So, what do you think made the difference?

GL: I think that part of it is about the timing of the research. The study that we modelled, which was done by Ian Edelstein for his PhD research in Khayelitsha, was specifically focused on a programme that was already established there. The key objective of his study (and later also ours) was to identify what impact an intervention run by Amandla EduFootball had on youth violence in the area. Amandla

EduFootball had already built their facility, and it was well known in the radius he was surveying. He was interested in the impact that it had on children who were taking the programme. In Gugulethu, where we were working, they had built the facility as part of the secondary school near Nyanga Junction Station, but I don't think it was that well known to the surrounding communities. We wanted to start the Gugulethu and Manenberg studies with baseline data collected prior to the intervention being run so that we could gauge awareness of Amandla EduFootball, as well as how many young people took the programmes at the Safe Hub, and how it affected their behaviour over time.

NM: It is also important to add that the accessibility of the facilities is not the same. If you just look at them, there is a fence around the fields in Gugulethu, whereas there is no fence in Khayelitsha. Anybody who passes through the area can easily go there, and so kids just get given balls and they can start to become involved. I think that this is one of the success stories of youth-focused projects that have been done in Khayelitsha because its open accessibility has meant that everyone 'owns' the project and are the ones that actually protect it. You cannot steal a ball from there, because everybody in the community knows how this project benefits them. On the other hand, at the Gugulethu facility you feel like you are in someone else's territory or their yard. And so the community feels that it is OK to steal from there because it is not yours. So there is also that ideological difference.

KM: What is so interesting about that last point is that programme design, which is not within your control at all, actually impacts on research design and execution by putting up barriers in much the same way. I suppose the number of challenges you have experienced in the project means that it requires real reflection on what conclusions can be reached from your data.

GL: One thing that this study has exposed is how difficult it is to do this particular type of research in a place like South Africa. Our experience has raised a number of worrying questions for us about studies that have been done before, particularly about the quality of data. Our study has shown just how difficult it actually is to get reliable data in large-scale randomised household surveys. We used two different service providers through this project and experienced very high levels of fraud on the part of the companies and/or unreliable fieldworkers – for example, fieldworkers were filling out questionnaires themselves, making up data and not following the sampling strategy. I think that this is a key issue that no one really wants to talk about. Studies publish and discuss their methods, but they gloss over the detail because this is really where the weak underbelly is. And the risk is that discussing it can undermine the findings.

KM: I suppose, then, that my last question may be an obvious one: would you do it again?

GL: I think we would do it differently – with the correct resources, the correct research strategy, and a lot more money, which would allow us to have the requisite infrastructure. The most successful longitudinal study in South Africa currently is the National Income Dynamics Study (NIDS), and they receive a large amount of funding from the Presidency, and have an entire organisation devoted to keeping the study running. They are also focusing specifically on households and not on individuals, and it is an ongoing project where they continue to incentivise participation. For us, the learning is that you can't do this on the cheap, and you also can't only fund the project partially – in other words, having enough money to only do a baseline, with the aim of getting additional funds for follow-up phases based on those results. So certainly it would mean a lot more planning, and if we did it ourselves, we would do the project ourselves completely. We would hire our own people, we would have them on proper contracts, and see the effort as a multi-year study where they are hired for the duration as far as possible so that the fieldworkers get to know the families as well. I think that's the only way of doing it. Trying to do it in the way we have done this project is highly risky and we encountered problems because of that. There's no cheap way of doing longitudinal studies – I think that's what we've really discovered out of this.

NM: You have to invest in making sure that your fieldworkers actually understand the data that they are generating. They should not only know about the findings, but they should understand what happens when the data is being processed, so that you reduce the likelihood of them compromising on data collection. The problem with hiring people in the way we did before is that they just want to do the job. And unfortunately research like this can't just be 'a job'. Little things that a fieldworker does may make or break an entire study, so you therefore need to invest in these people such that they actually understand their role and what it means for the larger project and its findings. We need to let them understand that disclosing any issues that may have compromised the data is as important as reporting that you have completed your tasks towards finishing the entire study. We need to incentivise fieldworkers to tell you what actually happened in the field, especially what went wrong. And if we do it again, we need to start from scratch again with planning, even if we were to go back to the same area. We almost need to forget some of the things we already know, so that we can avoid making assumptions through the process, and instead recheck everything from scratch.

GL: It's been a particularly valuable learning experience for our team. The learnings for us are that if we were to ever do it again we would need to raise the necessary money for it and have the necessary resources and support for the project.