The value of unrepresentative surveys:
collecting data on migrants through service provider NGOs

For presentation at the African Migrations Workshop, Rabat, Morocco
26-29 November 2008

Tara Polzer
Forced Migration Studies Programme
University of the Witwatersrand

Abstract

Conducting methodological defensible, logistically feasible and affordable large-scale national surveys of migrants is a serious challenge. Based on the experience of running a multi-year Migrant Rights Monitoring Project (MRMP) in South Africa, this paper outlines the pros and cons of working with and through NGOs which provide services to migrants (including asylum seekers, refugees and undocumented migrants) in order to conduct a national longitudinal survey on migrant access to basic public services. From a narrowly defined, formal methodological perspective, this access method clearly does not result in a sample which is representative of a total national population of migrants. Some of the resulting limitations are outlined in the paper, including urban bias, nationality bias, documentation bias, gender bias, and vulnerability bias, as well as effects relating to respondent duplication, strategic responses and intermediary quality and interests. The paper argues that these biases are problematic and we need to be aware of them, but that they are not significantly greater than for migration surveys which attempt to use more representative sampling methods. Furthermore, there are strong benefits of such an NGO-linked methodology. Recognising these benefits requires a wider reference point than formal methodology, namely the overall purpose of the research. Apart from making larger and more longitudinal surveys logistically and financially possible in the first place, such benefits include the formation of active and collaborative networks among organisations in the migrant rights sector; capacity building within this sector around research and the use and meaning of empirical data; and the direct integration of empirical data into local and national advocacy work.

Introduction: Surveys and Representativity

In most social science methodology textbooks, representativity = surveys and surveys = representativity. Not only are surveys seen as the quintessentially representative method (as contrasted with qualitative methods or experimental research designs); representativity is the key test of a survey's quality. A survey's fundamental purpose is to enable generalisability of findings to a wider population on the basis of statistically significant extrapolation from a representative sample. If this is the *raison d'être* of surveys in general, can a survey based on particular kinds of non-probability (and

therefore non-representative) sampling ever be justified? Can the findings from such a survey still lead to valuable outcomes?

In this paper, I argue that in the study of cross-border migration in contexts of dispersed (especially urban) self-settlement, non-representative surveys which are collected through 'service provider organisations'[1] are not only justified, but indeed have a range of process-benefits over other (attempts at) representative surveying of migrant populations. This is in a context where, as Bloch argues, the lack of an accurate sampling frame poses an "unassailable barrier to the textbook ideal type" of a representative survey (Bloch 2007:213). All surveys therefore have to creatively compromise and work with certain generalisability limitations and biases. I attempt to show how working with and through SPOs, while sharing many of the biases of other sampling methods, can be especially effective when generating survey data that is intended to inform policy advocacy, if the relationship between researchers and SPOs is explicitly collaborative, builds capacity and leads directly into advocacy work.

There are three elements to this argument, each addressing a different kind of methodological problem. The first problem is the most commonly written about, relating to the nature of the population being studied: self-settled migrants are notoriously difficult to sample accurately, and so working through service providers is a flawed access point, but one which is not necessarily more flawed than other access points (such as house-to-house residential sampling). The lack of a sampling frame justifies a sampling methodology which aims to approximate generalisability and reduce bias through volume and particularly through breadth of data rather than through the formal 'randomness' of the sampling method. Data breadth here refers to a kind of post-hoc stratified sampling, where the presence of certain previously identified groups (such as women, certain nationalities, undocumented, etc.) is monitored to ensure their inclusion in the overall sample. The final sample therefore cannot be used to make generalisations about the migrant population overall (e.g. what percentage of migrants in South Africa are Zimbabwean), but, like other stratified samples, it can be used to compare groups within the sample (e.g. are Zimbabweans significantly more likely to be unemployed than Congolese).

The need for volume and breadth brings us to the second problem, which is rarely discussed in academic methodology writing, but which is very familiar to most researchers in practice (especially in the global 'south'): the question of financial and logistical feasibility. Some research organisations are able to raise and dedicate the resources required for a national survey of migrants with thousands of respondents and tens of field workers, and such surveys are very valuable. In practice, however, such surveys are few and far between and many organisations with worthy research questions and agendas cannot afford them. Given limited resources, small research teams and inhospitable survey sites, how can one achieve the broadest and largest possible sample

[1] As described further in the case study description section, I use 'service provider organizations' (SPOs) to refer to NGOs and religious organizations which provide different kinds of services to cross-border migrants, including refugees, asylum seekers and, in some cases, documented and undocumented economic migrants.

of migrant respondents? I argue that working with and through existing organisations is a very cost-effective strategy and may be the only financially feasible option in some cases.

Third, and in my opinion most important, is the data use problem, which is in effect an ethical problem. What will the collected data be used for and who will use it? Although good research methodology textbooks always state that data collection techniques should be designed with a clear data use in mind (know your audience, etc.), there is rarely any discussion in these textbooks, or in the methodological literature, of data dissemination techniques and how these might relate to data collection. It is often unclear whether the academic researchers who collect social science data can or will directly transform their findings into advocacy outputs (apart from cases of government-commissioned research). This is particularly striking in migration and forced migration research, in spite of the oft-claimed advocacy goals of the research (Polzer 2007). This is fundamentally an ethical question because asking migrants or refugees (or any vulnerable people) about their needs creates an obligation for action on those needs (Turton 2002; Jacobsen and Landau 2003). By working with service provider organisations to collect the data, I argue, they not only gain skills in basic research techniques, but also develop a sense of ownership over the data, enabling them to use the findings more effectively in their own advocacy and intervention work and therefore spreading the likelihood that the data will be used, and the levels (local and national) and ways (grass-roots awareness raising through national policy influence) that it will be used.

This discussion therefore takes the relatively common practice of conducting refugee and migrant research through NGO or community-organisation intermediaries from being merely a convenience issue (as mentioned critically by Jacobsen and Landau 2003; Bloch 2007) to being a more reflected and conscious process in terms of the values of ‘action research’ (as described in more detail below).

I will make these arguments using the example of an ongoing national survey managed by the Wits University Forced Migration Studies Programme, South Africa. This survey is part of the Migrant Rights Monitoring Project (expected duration June 2007 – end 2009). The project does not only include data collection through SPOs, but adds questionnaires collected at governmental sites (Refugee Reception Offices) as well as a still-outstanding community-based survey. Being able to augment an SPO-based survey through such complimentary sampling methods is obviously ideal, but I argue that even a stand-alone SPO-based survey has its benefits, once the limitations are clearly understood.

In the rest of the paper I briefly summarise some of the recent methodological literature on surveying migrants, in particularly in Africa and South Africa, with a focus on common biases and limitations encountered in attempting large-scale surveying of self-settled migrant populations. I note some of the options suggested for addressing these limitations, and identify some key gaps in current methodological discussions. I then present a description of my case study: first a very brief note on migrant populations in South Africa and what makes them a difficult population to survey, and then an outline of the service-provider survey method I employed. This is followed by an analysis of the

biases introduced by this method, an evaluation of whether these biases are worse than other survey techniques, and of their possible impacts on the usefulness of the data. The limitations discussed include urban, nationality, documentation, gender and vulnerability biases, as well as respondent duplication, strategic responses and 'intermediary' quality control and interests. Finally, I outline the key benefits of the 'service-provider' survey approach, which are logistical and financial feasibility, the catalysation of active networks, research capacity building within the migrant rights SPO sector, and a more direct pathway from research to advocacy.

Surveying Migrants: Recent Scholarship

After a long phase of relative silence on methodological issues in forced migration research (Jacobsen and Landau 2003), methodological reflection has recently expanded dramatically (see JRS Special Edition 20(2) in 2007, for example). Since refugee and migration studies are multi-disciplinary fields, the role of large-scale quantitative studies is contested, although they are much more established in demography-centred migration studies than in legal and political-science centred refugee studies. While some see surveys as generally inappropriate for capturing what is really important about refugee and migrant experiences (Rodgers 2004), others see surveys, in all their imperfection, as still fulfilling important academic and, not least, strategic goals. In the latter camp, Jacobsen and Landau argue that large-scale, quantitative data on migrants is important for making well-founded policy recommendations, which cannot appropriately be supported by isolated and small-scale case studies (Jacobsen and Landau 2003). I argue here that large-scale, national quantitative data on migrant public service access is indeed necessary to address policy-makers in South Africa, and, perhaps counter-intuitively, that generating such data through SPO-based (and therefore not strictly representative) sampling is not only legitimate but desirable.

Surveying cross-border migrants in a representative fashion is well-known to face a series of common difficulties. Fawcet and Arnold enumerate four problems which all apply directly to the discussion below (Fawcett and Arnold 1987: 1531-2). These are the lack of a sampling frame; high costs of national and longitudinal studies; high non-response rates; and the likelihood of dishonest and strategic responses. Fawcet and Arnold also note the common tendency of researchers to use unrepresentative sampling methods to address the lack of a sampling frame, including organisation-based or snow-ball sampling, or sampling from areas of high concentration of out- of in-migration (ibid). SPO-based sampling to a significant extent addresses the problems of cost and non-response, as discussed below, while not exacerbating the problems of strategic responses and general non-generalisability, compared with other common strategies to deal with the lack of a sampling frame.

Alice Bloch (2007), in a recent review of methodological challenges in surveying refugees, adds to and elaborates on several issues raised by Fawcet and Arnold. These include the usefulness of a trusted intermediary between researcher and respondent to address issues of fear and suspicion (ibid:234, 236), and the dangers of gatekeeping and limited networks if respondents are identified by snowballing from a limited number of refugee organisations (ibid:235). Bloch also notes the importance of language in ensuring

accessibility to respondents and comparability of translated questionnaires across communities (ibid:240), which links with her discussion of appropriate modes of data collection (e.g. face-to-face, written self-completion, etc.). She concludes by emphasising the need for extensive exploratory work with the target community and flexibility during the field work process as necessary for enabling good research quality. As discussed further below, the process of working through SPOs over time addresses Bloch's concerns by having the benefit of providing a trusted intermediary without most of the gatekeeping and limited network dangers of community-based snowballing. Our survey was also designed to enable flexibility in allowing for either face-to-face interviewing for those with less literacy and questionnaire self-completion in five languages for those able to do so.

Additional issues raised by authors critically evaluating recent surveys of migrants and refugees in South Africa are the fundamental inadequacy of most spatially-defined random sampling frameworks in inner-city areas, and logistical issues of safety and response rates. These undermine standard attempts at spatially-defined representative sampling, even within relatively narrow pre-selected neighbourhoods and within pre-determined national quotas (Vigneswaran 2007). Singh et al present a highly involved spatially-defined sampling methodology which does claim to enable representative sampling (again with narrow neighbourhoods and nationality groups), but they also encounter extensive and expensive logistical difficulties in implementing this framework (Singh, Clark et al. 2008). While the creative efforts made by these researchers to approximate representative sampling methods are valuable, they nonetheless highlight the usefulness of considering non-spatially defined sampling frames.

The key concern of all these methodological discussions is the quality of the data collection process. None discuss what is to be done with the data once it has been collected, nor whether any link exists between modes of collection and later data use. These questions are central to debates on 'action research.' Some researchers see extensive consultation with 'communities' and migrant organisations before and during the research as mainly about improving the quality of the research findings (by enabling access, identifying where 'hidden' migrants are located, ensuring cross-cultural comparability of concepts and terms, etc.) (Parrado, McQuiston et al. 2005; Bloch 2007). I am arguing for a broader approach to consultation and collaboration, although without the more esoteric self-transformative elements of some 'action research' proponents, given that my focus is on incorporating service provider organisations into research and not on incorporating the migrants and refugees themselves. This approach includes seeing process as integral to achieving an outcome, and the values of being 'critical, evaluative, systematic, strategic, participatory, collegial, collaborative, self-reflective about practice, [and] empowering', rather than formalistic and formulaic, in research (Melrose 2001:161). These values are appropriate to research in a field like migration and refugee studies for their own sakes, but also because they directly contribute to improving outcomes for the people being studied.

The Case Study: the Migrant Rights Monitoring Project Public Service Access Survey

The appropriate method for monitoring migrant rights depends to a large extent on the context of migration in a country. South Africa is the regional economic hub as well as one of the most stable and prosperous countries on the continent, and so has been at the centre of a centuries-old regional labour migration system as well as attracting new flows of forced migrants from across the continent since the early 1990s (Wa Kabwe-Segatti and Landau 2008). South Africa has a policy of urban self-settlement for asylum seekers and refugees, rather than constraining them to camps, and economic migrants are also widely dispersed throughout the country. Important and large groups of migrants are undocumented, including many economic migrants from the region, most Zimbabwean migrants and refugees fleeing economic and political crisis since 2000, and significant numbers of other refugees who are not able to access the inefficient asylum application system (CoRMSA 2008). A methodological implication of this migration context is, as in many similar contexts, the lack of either clear spatial or bureaucratic sampling frames, logistical and language difficulties in accessing dispersed and diverse populations, and widespread distrust of researchers.

Because of the urban self-settlement policy, most basic welfare needs are expected to be met by public services (schools, clinics, etc.) and by the migrants themselves through the open market (rental accommodation, employment, etc.). There are therefore a wide range of non-governmental organisations that provide services to migrants, but mainly in an auxiliary capacity, without providing complete care. Types of service provider organisations include legal advice offices, basic welfare organisations, shelters, special-issue advocacy organisations (e.g. access to education or health), faith-based organisations, and refugee self-help organisations. There are widely differing levels of organisational formality and reach, with some working nationally and many limited to one locality. Mandates are also affected by whether the organisation is a formal UNHCR implementing partner, in which case they are supposed to only service recognised asylum seekers and refugees. Other kinds of organisation, for example faith-based groups, generally do not discriminate by legal status. There are some existing networks among these organisations, including city-based networks and a national network (under the Consortium for Refugees and Migrants in South Africa – CoRMSA), but apart from holding regular information sharing meetings, these networks do not have many shared activities.

It is in this migration and organisational context that the Migrant Rights Monitoring Project (MRMP) was developed by the Wits Forced Migration Studies Programme.[2] The MRMP is a multi-year research programme (July 2007 – end 2009, with possible extension) whose aim, as the name suggests, is to monitor the extent to which the legal rights of migrants and refugees in South Africa are being upheld in practice and to provide national data showing change over time in rights protection. The data collected is explicitly designed to inform advocacy activities towards the protection of migrant and

[2] The MRMP is, in turn, part of a larger migrant and refugee rights programme funded by Atlantic Philanthropies in South Africa. FMSP gratefully acknowledges the support from Atlantic Philanthropies.

refugee rights by other organisations in civil society, and to inform governmental action. The MRMP covers a wide variety of rights (including access to documentation, rights during arrest and deportation, and socio-economic rights), and we use a wide variety of methodologies to monitor these rights (including several different surveys, in-depth case studies and qualitative methods). However, this paper will focus only on the MRMP Public Service Access Survey (PSAS). Apart from covering basic demographic information about migrant respondents (sex, age, length of time in country, documentation, education level, etc.), the PSAS focuses on access to socio-economic rights in education, health care, housing, employment and social welfare.

To provide a sense of the survey's scope, here are some brief examples of the kinds of information produced by the PSAS (based on analysis of the 890 SPO questionnaires collected between July 2007 and February 2008):

- 85% of migrant respondents do not have school age children with them in South Africa, but 35% of those school age children who are in the country are not attending school.
- More than half of the respondents have not required health care since their arrival in South Africa, and 73% of those who have did not experience any problems accessing health care. Language problems and lack of documentation were the main kinds of problems experienced. Significantly more problems in accessing health care were reported in Durban than in Cape Town, Johannesburg or Pretoria.
- 70% of respondents live in privately rented flats, with 60% of these in subtenancy arrangements. The dominant accommodation challenge is overcrowding, which is significantly more common for undocumented than for documented migrants.
- Regarding employment, the well-educated (tertiary degrees) are no more likely to be working than the less educated (primary education), and police harassment was the most commonly noted concern among those who were working.

Between July 2007 and June 2008, the project had collected 1290 questionnaires through partner SPOs around the country and data collection is ongoing. The PSAS is being implemented through ten partner service provider organisations (SPOs) in the four main urban areas of South Africa (Johannesburg, Pretoria, Cape Town and Durban). [3] The SPO selection was intended to be as broad as possible for each city, but with an initial focus on established and institutionalised NGOs over smaller community-based and refugee-self-help organisations (although one partner is a Zimbabwean self-help organisation in Johannesburg). This is because partner organisations were intended to use their own, existing staff resources to contribute to the project, and because we were worried about the internal politics of the fragmented refugee organisations. Furthermore, working through refugee organisations would have biased the sample too strongly towards documented and political migrants, while we were also interested in undocumented and economic migrants (although this bias remained in place in working through established NGOs as well, as discussed further below).

[3] These SPOs are the African Disabled Refugee Organisation, Bechet School, Cape Town Refugee Centre, CARE (Johannesburg), Excelsior Empowerment Centre, Lawyers for Human Rights, Mennonite Central Committee (recently renamed Refugee Social Services), Mthwakazi Arts and Culture, South Africa Red Cross Society and the Scalabrini Centre.

Data collection and respondent sampling has had several characteristics which have been important to the success of the project so far: a) the continuous, longitudinal collection model, b) maximum coverage but flexible sampling of SPO clients, and c) ease of engagement with the instrument by SPOs and respondents. As stated above, these measures were all intended to enable volume and breadth of data collection, rather than comparability or formal 'randomness' of respondent sampling.

The data collection has been continuous, reaching each new SPO client as they arrive into the service provider programmes, rather than being concentrated in a short time period. On the one hand, this gives the data a continuous longitudinal element, through which changes in migrant experiences can be traced month by month over time in aggregate. This gives the survey (especially after its 2 year running time) a significant advantage over other surveys which have either been once-off cross sections (CASE 2003) or repeated cross sections with significant breaks in between (such as the FMSP 'African Cities' Johannesburg Study in 2003 and 2006). As importantly, the continuous process has meant that SPOs could fit the data collection into their ongoing work rather than taking time and resources out of their programmes to focus on data collection for an intensive period of time.

This leads into the maximum coverage but flexible sampling method adopted. Each SPO was requested to include as many of their new clients as possible in the survey, preferably all new clients, but it was up to each organisation to decide on how this would be accomplished. In some cases, clients would complete the questionnaire while waiting to see legal counsellors or social workers, while in others the questionnaire would be built into regular rights-awareness training workshops. In other organisations, it was included in the exit-interview after a 6-month welcoming programme for new migrants, or into English classes for French-speakers. In practice, there were some problems with continuity of commitment by SPO staff (partly due to insufficiently regular check-ups by myself), so that actual coverage of non-citizen clients was not close to complete. This introduces an additional element of uncertainty into the sampling (as discussed further below on quality maintenance). However, given the emphasis on volume and breadth in the overall sampling concept, this does not invalidate the entire approach, since it was understood from the beginning that regularity and comparability of the sub-samples making up the surveyed whole would not be possible in any case.

The comparability of responses was brought in through the use of a relatively simple and accessible questionnaire of closed, pre-coded questions, used by all SPOs. This questionnaire was translated into English, French, Swahili, Somali and Portuguese.[4] Complementing the flexibility premise above, the questionnaire is designed to allow for self-completion (on the SPO premise), or for assistance/interviewing by an SPO staff member. To facilitate SPO assistance to respondents without requiring staff to speak all the relevant languages, the non-English versions of the questionnaire also include the English original for each question. This still created problems for respondents who are

[4] The translations were confirmed by an independent second translator familiar with the project to enable identification and discussion of possible misunderstandings, but were not formally blindly back-translated.

not literate in their home languages (or whose home languages were not included in the translations) if no SPO staff could speak their language; an issue, as elsewhere, which was particularly noticeable for Somali women, as reported by SPOs. We have recorded whether the questionnaires were self-completed or assisted and so will be able to evaluate whether there were any significant differences in the quality or nature of the data using these different modes.

Because of the relatively long time-frame of the 'data collection phase' of the research, and the commencement of 'data analysis' and 'advocacy' while data was still being collected, some level of iterative adaptation of the research process has been possible. For example, the original group of participating NGOs expanded (and is still expanding), and there has been a shift from initially not providing any funding to partner organisations to now providing some organisations with small monthly stipends to support data collection interns or volunteers. This change, while increasing the volume and predictability of monthly data collection, may also impact on data quality issues (discussed below) by creating new incentives for falsifying data. The impact of this change remains to be seen. As above, we will look at the data provided by the newly paid organisations to see whether there are any data regularities or irregularities that may point to falsified data.

The long time frame and continuous, iterative engagement with SPO partners enables the most important element of this survey methodology: namely the partnership and capacity building element discussed further in the last section below. The survey instrument was designed after consultation visits with all partners and a draft was then work-shopped with them again. Each SPO was visited individually and trained in the purpose and use of the questionnaire. Each SPO is also called every two months to provide feedback on the data collection process. Completed questionnaires are posted by SPOs to the Wits Forced Migration Studies Programme as soon as a stack is collected, and so data entry is continuous. So far, the entered data has been analysed twice (in January 2008 with 317 completed surveys and in June 2008 with 890 completed) and findings shared with all SPOs in a data summary report. These reports include the raw percentages of answers to each question as well as narrative analysis of the key findings and basic advocacy pointers arising from the findings. The initial January 2008 data summary was again work-shopped with all SPOs (and other organisations not formally part of the partnership) to ensure that the format was understood and useful. All SPOs are encouraged to request specific analysis of the data to suit their specific advocacy or programming needs (e.g. pulling out comparative data on education access across cities, or comparing Zimbabwean access to education with other nationalities, etc.). This offer has so far been used only by two organisations, but in those cases has successfully contributed to local advocacy and awareness-raising campaigns.

Limitations of Surveying through Service Providers

Surveying migrants through SPOs in South Africa, as described above, brings certain biases with it, many of which are likely to apply in other country contexts as well where there are self-settled and diverse (rather than camp-based) migrant populations who are served by NGOs. These include urban, nationality, documentation, gender and vulnerability biases. Other data problems are respondent duplication, strategic responses

and the quality of SPO staff as 'researchers', as well as their potential interest in constructing certain outcomes in the data. I will comment briefly on each of these issues and note the extent to which they pose greater or similar liabilities in data quality compared with some other recent South African migrant surveys.

Urban bias

Working through SPOs in South Africa introduces two kinds of urban bias: urban versus small town and rural bias, and inner city versus township and informal settlement bias. Virtually all migrant-oriented SPOs are based in the main metropolises in South Africa (Johannesburg, Pretoria, Cape Town and Durban). These are the same cities where the government's Refugee Reception Offices are located (apart from Port Elisabeth, where there is an RRO but relatively few established migrant SPOs). While there are clearly large concentrations of migrants in these cities, there are also important and large non-citizen communities in the rural border areas (especially bordering Mozambique and Zimbabwe), and largely unstudied but significant non-citizen populations in smaller towns and secondary cities around the country (CoRMSA 2008).

The urban/non-urban distribution of migrants is of course not random: there are important differences between the average characteristics of urban migrants (more refugees and asylum seekers, more educated, often from urban backgrounds, some women and families), small-town migrants (mainly single, young, entrepreneurial men from specific national communities such as Somalis, Ethiopians, Pakistanis who may or may not have asylum documentation), and rural border area migrants (mixture of long-term residents and circular labour migrants, less education, often from rural backgrounds, very few with asylum and refugee documentation). Furthermore, the different contexts lead to very different experiences for non-citizens in terms of access to basic services and relations with host communities.

Regarding the second kind of urban bias, most SPOs are based in or close to inner cities. This means that they are well-located to serve migrants resident there, but not necessarily easily accessible to residents of townships and informal settlements more removed from the city core. The Excelsior Centre in Cape Town is an exception to this in our set of partner SPOs, as is the Mthwakazi Arts and Culture organization which does regular outreach to Zimbabweans in townships and informal settlements around Johannesburg. Some of the legal organizations have clients who travel from outer-city areas and even other towns to seek assistance. In spite of these exceptions, people living outside the inner city are relatively underserviced by SPOs, and are therefore also not reached by our survey. As above, this is important because the demographic and experiential characteristics of inner-city and informal settlement-based migrants are likely to be significantly different by nationality, education level, gender distribution, employment type, etc. The 10% of MRMP respondents to date who do report living in informal settlements are significantly more likely to be Zimbabwean, undocumented, less educated, and unemployed than those living in inner-city accommodation.

Most previous surveys of migrants in South Africa share these forms of urban bias by either explicitly limiting the geographical areas of focus to specific neighbourhoods in

inner-city Johannesburg (Vigneswaran 2007; Singh, Clark et al. 2008) or Johannesburg and other cities (McDonald, Mashile et al. 1999), or surveying through urban-based Refugee Reception Offices and urban-based refugee communities (CASE 2003). Exceptions are the national migration survey carried out by HSRC in 2001-2 which included rural and peri-urban areas (Kok, Van Zyl et al. 2006).

Nationality bias

Different nationalities have different likelihoods of using formal service providers. For example, Somalis, Ethiopians, Eritreans, and Mozambicans (who, after Zimbabweans, are the largest foreign nationality in the country but only make up 0.3% of the survey respondents) do not use the mainstream migrant rights SPOs very often, and certainly less than their respective percentages in the overall non-citizen population in South Africa. Asian migrants from Pakistan, Bangladesh and China also do not make use of migrant NGOs. In contrast, Congolese and Zimbabweans are well-represented among SPO clients, but not according to their respective prevalence in the overall population. Congolese make up 37.5% of the survey respondents, followed by 29.3% Zimbabweans, even though there are clearly many more Zimbabweans in the country overall than Congolese. Nationality bias in this survey is exacerbated by the inclusion of some nationality-based SPOs. For example, in the initial group of SPO partners, our only Johannesburg-based partner was an organization catering mainly to Zimbabweans. This meant that the Johannesburg data was not comparable with other cities, since it only reflected the experience of one nationality, and one with a significantly different profile than other nationalities. It was possible, of course, to compare the experience of Zimbabweans across cities, but we then added additional SPOs in Johannesburg and now have a more balanced sample.

Again, other migrant surveys (as well as qualitative studies) in South Africa commonly address problems of cost and research focus by pre-limiting the nationalities they sample. The 2006 African Cities Project in Johannesburg only surveyed Somalis, Congolese and Mozambicans (and South Africans) (Vigneswaran 2007), while Singh et al. focused exclusively on Somalis, Congolese, Zimbabweans (and South Africans) (Singh, Clark et al. 2008). The Southern African Migration Project selected Basotho, Mozambicans, Zimbabweans, Malawians, Nigerians and 'Francophone Africans' as target groups, with other nationalities included in small numbers under sampling rubrics such as 'traders' (McDonald, Mashile et al. 1999). The overall effect of such inbuilt nationality bias is that certain nationalities are almost never included in surveys (such as Ethiopians and Eritreans, or East African migrants). Apart from specific targeting of certain nationalities by surveyors, ethnographic research in Johannesburg confirms that certain nationalities – particularly Mozambicans – are particularly keen to remain 'hidden' from official or public eyes (Madsen 2004; Vidal forthcoming), showing their general resistance to surveying, and not only to SPO-based sampling.

Documentation bias

Documentation bias is very common in migration studies around the world. Our SPO survey had two kinds of documentation bias: self-selection and mandate-based selection. Undocumented migrants are often afraid to ask for help, do not trust institutions or do not

think they have any rights which SPOs could assist them to claim. In addition to this self-selection, some organizations, especially those who are UNHCR implementing partners, have mandates to assist only documented asylum seekers and refugees. The result in our survey is that only 21.1% of respondents reported being undocumented (which might be an undercount, since respondents may have claimed documentation they did not really have – see below on strategic responses). In the overall migrant population in South Africa, the undocumented are in the majority, compared with asylum seekers and refugees. But given the sensitivity of documentation issues and that only some SPOs actively assist the undocumented, including 21% of undocumented respondents is actually a surprisingly high percentage.

The common problem of estimating wider documentation levels from migration surveys is illustrated, as one of many examples, by the SAMP survey which sampled its 501 respondents through a version of 'snowball' sampling starting from migrant community insiders. On the basis of this non-representative sampling system, the authors then state that 93% of respondents had some kind of official documentation. While they claim that this disproves the stereotype of migrants in South Africa not having documentation (McDonald, Mashile et al. 1999:174), it is just as likely to be a sample self-selection and mediator-selection effect. The 2003 National Refugee Baseline Survey explicitly incorporated documentation-bias by only interviewing documented asylum seekers and refugees (CASE 2003). Door-to-door surveys such as the African Cities Project have high refusal rates, which are likely to be biased by documentation status as well.

Gender bias

Woman are generally more likely to seek welfare help than men: this is a global phenomenon. Overall, the migrant and refugee population in South Africa has relatively few women. Although we might expect a higher percentage of women to be accessing SPOs than their percentage of the overall population, we cannot clearly evaluate this, due to the lack of a baseline. This bias may be less than expected, however, if we look at results from the service access survey that we conducted at the Refugee Reception Offices (e.g. not pre-selected for persons who were already accessing SPO assistance)[5]: there was no significant difference between male and female respondents on whether they reported ever having received welfare assistance from an NGO.

As above, most other surveys either explicitly stratified their samples to include a certain gender balance or reported various gender-related biases (such as women being more likely to be at home during the day when surveyors were active) (Vigneswaran 2007). Singh et al went to great lengths to ensure random selection of respondents within households while taking into account cultural sensitivities about women only being interviewed by other women, and so were likely to have overcome most gender bias in their sample (Singh, Clark et al. 2008).

Vulnerability bias

[5] The Pretoria (Marabastad) survey was conducted by trained field workers over the space of two weeks in November 2007 (N=364), and the Durban survey was conducted in February 2008 (N=300) using the same questionnaire as the SPOs.

While urban, documentation and gender bias are well-known and commonly discussed biases in migration surveys in general, vulnerability bias is a key bias when working through SPOs and has two elements: the need for assistance, and the ability to seek assistance. SPOs clearly only serve those people who need some kind of help, thereby excluding those from the sample who have no immediate legal, welfare or counseling needs. Non-citizens with enough resources (financial, social, informational) to look after themselves are therefore invisible. This is a serious concern if the goal is to achieve an overall picture of migrant life in SA, since much of the positive and affirming information is therefore missing. From an intervention and advocacy perspective, however, what is particularly concerning is that people who have needs may still be excluded from this survey if they do not have enough resources (financial, social, informational) to reach SPOs in search of assistance. When comparing results from the RRO survey with the SPO survey, we see that respondents at SPOs are significantly more likely to have at least a completed secondary education than those at the RROs.[6] This suggests that the least educated either have fewer assistance needs (which seems unlikely) or that they face barriers in accessing SPO assistance.

Non-SPO-based surveys might be assumed to have less vulnerability biases, but often levels of vulnerability are hidden in other biases, such as location and time-of-day biases. More economically successful migrants might be less likely to live in the oft-sampled 'migrant neighbourhoods' in the inner city, while working migrants are less likely to be at home in day-time house-to-house surveying.

Respondent duplication

In addition to these common respondent selection biases, there are several further data quality issues which relate to the actions of respondents and research 'intermediaries' once the respondents have been identified. One of these is respondent duplication. The reliability of surveys depends on each respondent only being recorded once. This was identified as a potential problem early on in consultations with SPOs, who are very aware of the fact that migrants often go to several SPOs in a city for different kinds of assistance. We therefore specifically included a note at the top and front of the questionnaire asking respondents not to fill in the questionnaire again if they had already done so elsewhere. This was also clearly articulated in SPO training regarding how to introduce the questionnaire to respondents. However, it is possible that some respondents will have overlooked this, or ignored it on purpose, wanting to pretend that they had not already sought assistance from another SPO. In a worst case scenario, respondent duplication could make up a large percentage of responses in each city. While it is possible to check the data for identical or very similar responses, it is possible that repeat respondents might change some of the responses they provide from one place to the next. Since the data 'intermediaries' (e.g. SPO staff) are different at each location, in contrast to 'normal' survey field workers who move from location to location, this data quality danger is also unlikely to be uncovered 'on the ground' by recognising repeaters.

Strategic responses

[6] The effect remains highly significant even when taking differences of nationality and sex into account.

Similarly to respondent duplication, strategic responses are probably, due to the likelihood of respondents perceiving a connection between the answers they give on the survey and the kinds of services they are hoping to receive at the SPO. The questionnaire introduction clearly states that it is separate from the SPO at which it is being completed, and that answers will in no way influence the services respondents get from the SPO. It is nonetheless possible, indeed likely, that respondents feel that their responses will in some way influence immediate or future services (in spite of the anonymity of the questionnaire) and so might misrepresent potentially damaging information (such as claiming to have documentation when they do not) or exaggerate needs (such as the kind of accommodation they are in or employment needs). The same kinds of strategic reporting bias are likely in other kinds of migration or social service related surveys as well (Bloch 2007:242). The actual results of our survey, however, suggest that this is not a very large bias. Very few respondents reported that they were homeless, for example, with most stating they are in rental accommodation, even though access to accommodation is commonly identified as a key need and desire of migrants. Such strategic reporting might also have been reduced by the fact that there were virtually no questions about immediate food, clothes or money needs, with welfare-handout-related questions only dealing with whether such services had ever been received in the past (although past assistance may have been underreported in order to justify claims for future assistance).

Intermediary characteristics: mandate and quality

Researcher-effects are omnipresent in social science research, but 'textbook' quantitative surveys are intended to minimise these effects with the aim of generating 'objective' results allowing for a comparison of 'real' differences between respondents (Neuman 2000:69). When working through SPOs, there are two kinds of intermediary effects that can impact on the data. One is at the organisational level, as already mentioned above, where different mandates and focus areas can produce nationality, documentation and vulnerability effects (e.g. people coming to a legal organisation are likely to have different kinds of problems to those coming to a faith-based basic welfare organisation).

The second intermediary effect lies at the level of the individuals within the SPOs who assist respondents with the questionnaire and lies in the question of quality control. Since the implementation of the survey is, so to say, radically decentralised in an SPO-based model, the overall research coordinator has little ability to carry out immediate quality control by conducting spot checks or repeat interviews, or by getting to know field staff well enough to judge whether they are filling in questionnaires on their own, or similar 'cheating.' Everyday quality control is therefore left to the individual SPO's management, where it is likely to get little if any attention, given that the management does not get any personal or organisational remuneration for contributing to the survey. Such quality control is extremely difficult even in professionally managed large-scale 'representative' surveys, as evidenced by the evaluation of the 2001-2002 HSRC Migration Survey which found extensive 'cheating' by field workers (Van Zyl 2006), and so is by no means a liability unique to decentralised SPO-based surveying. It is theoretically imaginable that an SPO or its staff members might try to systematically adjust the questionnaire responses to fulfil a particular political agenda, although this

might be identified through careful analysis of the resulting data to identify any suspicious differences or patterns in responses between SPOs (for which there was no evidence in our survey). A much more common problem is simply wastage, where questionnaires are incomplete or wrongly or unclearly filled in, and then cannot be followed up once they have been sent in to the central collection point.

Impacts of these limitations on data use

Since the stated purpose of the Migrant Rights Monitoring Project is to produce data that is useful for advocacy purposes, what impacts do these biases and data limitations have on the usefulness of the survey findings? Clearly, as with other non-representative samples, an SPO-based survey can not be used to extrapolate or estimate various widely desired numbers, as a fully random national sample might. It cannot tell us the overall number of non-citizens in the country; the overall number of any national group and relative percentages of different nationalities in relation to each other; the overall gender-breakdown of non-citizens, or gender-breakdown within any national group; the overall documentation breakdown of non-citizens, or documentation breakdown within any national group; or overall levels of vulnerability and need among non-citizens (e.g. how many foreign children are in the entire country and how many of them are not in school), or the vulnerability profile within any national group.

However, there are many valuable kinds of information which the survey can nonetheless provide. These include a country-wide nationality, gender, documentation and vulnerability profile of those non-citizens who access SPO assistance. This can then be compared with other sources of data (such as surveys conducted at the Refugee Reception Offices, or conducted door-to-door in migrant residential areas in Johannesburg) to identify differences in profile and therefore who is not accessing SPO assistance. For improving SPO services and outreach, this is very important. Also possible are comparisons between profiles and vulnerability levels of those national groups who access SPOs. Such an analysis clearly shows that Zimbabweans are more vulnerable than other groups in terms of documentation and accommodation access, but not in terms of health care access, for example. It is also possible to make comparisons between cities on migrant profiles, especially with the use of regression analysis to ensure the exclusion of confounding factors (such as different nationality profiles by city).

One key drawback of SPO-based sampling is that it might impact adversely on government advocacy if policy-makers perceive the survey methodology to be illegitimate. This is a possible danger if officials, on the one hand, have enough methodological knowledge to ask how the data was collected and, on the other hand, do not have enough methodological knowledge about studying migration to know that virtually all methodologies are similarly flawed. ‘Representative sampling’ is a powerful signal phrase for policy-makers, whether or not actual representative sampling is achievable.

Benefits of Surveying through Service Providers

While SPO-based sampling might not have the same mystique (and therefore assumed policy legitimacy) as complex, multistage approximations of random sampling (which

are also useful and legitimate exercises, of course), it has several important benefits which can directly increase its policy impact. These relate to the 'do-ability' of the research in the first place, but mainly to the level of involvement in the overall research design by the same organisations who are intended to make use of the data, thereby shortening the distance between research and advocacy, and broadening the application of research findings. As noted in the introduction, this is an important ethical imperative.

Financial and Logistical Feasibility

A key barrier to conducting large-scale quantitative studies of migrants in southern countries is often that they are too expensive. An SPO-based survey requires paying an experienced researcher to design and pilot the questionnaire, build or expand on a network of SPOs (if it does not already exist for other reasons), train and maintain contact with the partners, analyse the data, and train partners in how to use the data. Data entry, once it has been collected monthly or sent in to a central place by the partners, can be outsourced or done by students. The major cost and logistical difficulty of a 'normal' survey falls away, however, which is hiring, paying and managing fieldworkers, and transporting them around the country. Even if small stipends are made available to SPO-based interns or staff members, this personnel expenditure has the added impact of sustainably strengthening the capacity of partner SPOs, and therefore the ability of the data to not only be collected but also to be used effectively.

SPO surveying also deals with other key logistical challenges commonly experienced during house-to-house research in (South) African cities, including researcher safety (Vigneswaran 2007). SPO-based surveying also greatly reduces the non-response rate, which is high in house-to-house and street-based surveying of migrants. A national randomly selected household survey on migration conducted by HSRC in 2001-2002 (N=4000 households) had a 43% non-response rate from cross-border migrants (Van Zyl 2006:148). Response rates are higher in SPOs because the respondent is in a place they already trust, to some extent, to be on 'their side', and because the person, by coming to that place, often has time on their hands already and does not have other work to do. This does not mean that there are no refusals; partner SPOs report outright refusals or partial refusals to complete the questionnaire especially when they attempt to use it during outreach activities with new clients.[7]

Active Networks

As mentioned above, there are existing networks of migrant and refugee rights SPOs in South Africa, including regional (city-based) networks and a national network (CoRMSA). These networks formed the basis for approaching SPOs to partner on the survey, although not all network partners are survey partners and some survey partners are not part of the formal networks.

The existing networks are generally used for basic information exchange and to some extent for coordinating advocacy campaigns, but there are very rarely joint activities over any period of time among network partners (except on a bilateral basis). The survey has been a practical and ongoing joint activity, which directly links the organisations in a

[7] Comments made during feedback session 26/09/07.

shared endeavour beyond their general shared interest in migrant rights. This is especially important for the smaller SPOs whose work is usually limited to their specific local clientele and who therefore appreciate the feeling that their work and experience is feeding directly into a larger, national project. Activating a network through a shared activity such as a survey can have several side effects, such as encouraging other kinds of regular information exchange within the network, catalysing bilateral collaborations, and generally energising local outreach and advocacy efforts, as described below.

Research Capacity Building for Evidence-Based Service Provision and Advocacy
Most migrant rights SPOs in South Africa do not have research experience. There are some large SPOs with extensive experience, of course, but often this experience is limited to specific people in the organisations or to policy research rather than client-based research. Furthermore, most SPOs, including the large and established ones, are focussed on implementing programming rather than conducting basic profiling research on their own clients or the wider pool of possible clients, and their needs. The MRMP, by working with and through SPOs, therefore fulfils a triple function: a) using SPOs to access migrants for information about the wider needs of non-citizens in South African, b) profiling the SPOs current and potential clientele to enable better internal management and sector-wide service planning, and c) building capacity within individual SPOs and the sector at large in conducting and using research as both an advocacy and a management tool.

Just as the SPO-survey is an imperfect but nonetheless useful instrument for the aim of gathering overall data on migrants in South Africa, so it is also imperfect and useful as a management tool. Due to wanting to prevent respondent duplication, for example, no SPO is able to consistently survey *all* their clients, even if they wanted to make the requisite staff effort. The questionnaire is also not designed to evaluate existing SPO services, but rather to establish broader migrant needs and the extent to which rights are met by public institutions. Nonetheless, several of the larger NGOs have recognised the management value sufficiently to build the MRMP questionnaire into their internal client tracking processes, enabling them to more effectively compare and link their own work with that of other organisations. Other partners now use the survey data about their own clients, and the aggregate data for their city, in internal and network-wide strategic planning exercises to identify key areas for intervention.

Finally, SPOs can benefit from learning more about research practice so that they are better able to plan and conduct their own research (even if small-scale) on issues affecting their clients, and so that they can use research conducted by others more effectively in informing their local work. The collaborative MRMP survey process assists in building this capacity through formal training sessions in using the questionnaires and in using the summarised data outputs, but also through the more informal learning experience of regularly using the questionnaire and therefore having an example of a research instrument that they become familiar with over time. For many SPOs, a key benefit is overcoming the fear and mystique of conducting research by seeing it as something they can be part of themselves. The latter learning experience has been

repeatedly stated by the smaller partner organisations in feedback discussions and trainings.

From Research to Advocacy

I come to the most important point of this paper right at the end. This is that the SPO-based survey process can contribute to narrowing the common gap between research process, research findings and the use of those findings in advocacy and implementation planning.

This gap has many elements. Often, research is conducted by different people than those who can use it for advocacy. Researchers, advocates and implementers are usually not only in different institutions, but act at completely different levels: local organisations are assumed not to be able to use national data and national organisations are assumed not to need or be able to use local data. Linking local organisations and the data they can collect in a nationally comparative study overcomes these divisions between institutions and levels of action and analysis. Secondly, advocates and implementers often cannot use research findings because they do not know about them. This is especially the case for local organisations, who may not hear about research that is conducted in other (but similar) locations or nationally, and if they do hear about it they cannot imagine how they can use it locally. There is also often no way for them to contact the researchers to be able to adapt the findings to their local needs. Participating directly in data collection gives such organisations ownership over the information and allows them to think about how to use it on a regular basis, as it becomes part of their own internal planning and performance management.

Finally, from the perspective of national professional advocacy organisations (such as CoRMSA in the case of South Africa), while they may have access to national or local research findings, they often miss opportunities for using data effectively because they are not present on the ground in different regions/cities. Local SPOs know more about when a local decision is about to be made about health care access or housing policy or street traders, for example, where research data could be used to good effect in influencing that policy. Similarly, local organisations do not have the capacity or knowledge to bring their local knowledge into national policy debates unless they are linked through such a collective survey.

Conclusion

This case study of the Migrant Rights Monitoring Project's Public Service Access Survey aims to illustrate two main points concerning methodological approaches to surveying migrants. First, migrant and refugee service providers should not only be seen as a 'bad practice' convenience solution to survey sampling, but as potentially valuable research collaborators. There are important representativity challenges and biases of working through service provider organisations, but these are also present in many other common sampling methods. Second, and more generally, we must incorporate considerations of data use directly into how we plan our data collection, and these considerations should not be limited to high-level policy makers' preferences for 'representative' data. Data

users, including effective advocates and implementers, should be seen at various levels, from the local to the national, and the design of survey processes can reflect this as well.

References

Bloch, A. (2007). "Methodological Challenges for National and Multi-sited Comparative Survey Research." Journal of Refugee Studies **20**(2): 230-247.

CASE (2003). National Refugee Baseline Survey: Final Report. Johannesburg, Community Agency for Social Enquiry.

CoRMSA (2008). Protecting Refugees, Asylum Seekers and Immigrants in South Africa. Johannesburg, Consortium for Refugees and Migrants in South Africa.

Fawcett, J. T. and F. Arnold (1987). "The Role of Surveys in the Study of International Migration: An Appraisal." International Migration Review **21**(4, Special Issue: Measuring International Migration: Theory and Practice): 11523-1540.

Jacobsen, K. and L. B. Landau (2003). "The dual imperative in refugee research: some methodological and ethical considerations in social science research on forced migration." Disasters **27**(3): 185-206.

Kok, P., J. Van Zyl, et al. (2006). The history and methology of the HSRC Surveys. Migration in South and Southern Africa; Dynamics and Determinants. P. Kok, D. Gelderblom, J. Oucho and J. Van Zyl. Cape Town, HSRC Publishers**:** 292-306.

Madsen, M. L. (2004). "Living for Home: Policing Immorality among Undocumented Migrants in Johannesburg." African Studies **63**(2): 173-192.

McDonald, D. A., L. Mashile, et al. (1999). "The Lives and Times of African Migrants and Immigrants in Post-Apartheid South Africa." SAMP Migration Policy Series **No. 13**.

Melrose, M. J. (2001). "Maximizing the Rigor of Action Research: Why Would You Want To? How Could You?" Field Methods **13**(2): 160–180.

Neuman, W. L. (2000). Social Research Methods: Qualitative and Quantitative Approaches. Boston, Allyn and Bacon.

Parrado, E. A., C. McQuiston, et al. (2005). "Participatory Survey Research: Integrating Community Collaboration and Quantitative Methods for the Study of Gender and HIV Risks Among Hispanic Migrants." Sociological Methods Research **34**(2): 204-239.

Polzer, T. (2007). Disseminating research findings in migration studies: methodological considerations. Johannesburg, Forced Migration Studies Programme.

Rodgers, G. (2004). "'Hanging Out' with Forced Migrants; methodological and ethical challenges." Forced Migration Review **21**.

Singh, G., B. Clark, et al. (2008). "Creating a frame: Random sampling in a non-homogeneously distributed urban migrant populations." DRAFT

Turton, D. (2002). Forced Displacement and the Nation-State. J. Robinson. Oxford, Oxford University Press.

Van Zyl, J. (2006). Evaluating the 2001-02 HSRC Migration Survey. Migration in South and Southern Africa; Dynamics and Determinants. P. Kok, D. Gelderblom, J. Oucho and J. Van Zyl. Cape Town, HSRC Publishers**:** 147-170.

Vidal, D. (forthcoming). "Living in, out of and between two cities: The migrants from Maputo in Johannesburg." Urban Forum.

Vigneswaran, D. (2007). "Lost in Space: Residential Sampling and Johannesburg's Forced Migrants." FMSP Migration Methods and Field Notes **4**.
Wa Kabwe-Segatti, A. and L. B. Landau, Eds. (2008). Migration in post-apartheid South Africa - Challenges and questions to policy makers. Paris, Agence Française de Développement.