Collecting data from migrants in Ghana:
Lessons learned using respondent-driven
sampling

Samantha R. Lattof

Contents

Collecting data from migrants in Ghana:
Lessons learned using respondent-driven sampling

Samantha R. Lattof[1]

Abstract

BACKGROUND
Policymakers and program implementers require high-quality data on migrants and migration in low- and middle-income countries (LMIC); however, a shortage of high-quality data exists in these settings. Sampling migrant populations requires better techniques. Respondent-driven sampling (RDS) may be one such solution.

OBJECTIVE
Using Ghana as a case study, the objectives of this paper are to: 1) assess RDS recruitment productivity, network size, and ties of internal migrants; 2) test for homophily; and 3) detail the successes of and challenges to implementing RDS in Ghana and how these lessons can be applied to migrant populations in other LMIC settings.

METHODS
This study used RDS to sample 625 rural–urban female migrants working as market porters (*kayayei*) in Accra, Ghana.

RESULTS
This study generated the most comprehensive data set on kayayei to date. Network size increases as participants become more educated and migrate more often to Accra. Ethnic group membership is an important determinant of recruitment, with certain groups preferring to recruit from within. Employing members of the kayayei population to collect data built crucial trust.

CONCLUSIONS
Whilst RDS is not a one-size-fits-all solution for sampling hard-to-reach migrants in LMIC, it can be a powerful tool to uncover and to recruit hard-to-reach migrant populations. In countries with multiple ethnolinguistic groups, recruiting a migrant population with greater ethnolinguistic overlap may facilitate quicker equilibrium.

---

[1] Department of Social Policy, London School of Economics and Political Science, UK.
Email: S.Lattof@lse.ac.uk.

CONTRIBUTION
This study expands the evidence base on use of RDS among migrant populations in LMIC and provides lessons learned to assist other researchers implementing RDS in LMIC settings.

1. Background

The 2030 Agenda for Sustainable Development marks the first time that the global development framework includes migration. Migration's relevance and contribution to development is reflected throughout the Agenda in goals and targets such as protecting labour rights of migrant workers, ending human trafficking, and implementing planned and well-managed migration policies (United Nations 2015). Monitoring progress of these goals demands high-quality, timely data on migrants and migration. Population registers or administrative systems can generate migration data; however, this data is primarily collected within Europe (UNDESA 2013). High-quality, timely data on migrants and migration in low- and middle-income countries (LMIC) is sparse. In LMIC, migration data comes primarily from decennial population censuses and ad hoc population surveys that are expensive to implement. Ad hoc subnational surveys and research on migration tend to be localised and small-scale; data are rarely representative or generalizable.

To better understand migration and migrants' lives in LMIC, researchers require better techniques to reach and sample migrant populations. Techniques must simultaneously minimize potential sources of bias and be resource (i.e., time, money, labour) efficient. Health and Demographic Surveillance Systems can be expensive to operate and may struggle to provide timely data (Ye et al. 2012). Limited resources may prevent the development of sampling frames, resulting in challenges to obtaining representative samples of highly mobile populations. Contextual factors and available tools for data collection may further impede researchers' efforts to survey migrants. Xenophobia, for instance, can lead to migrants being denied or deterred from health facilities, biasing facility-based surveys (Crush and Tawodzera 2014). Respondent-driven sampling (RDS), a technique originally developed to sample hard-to-reach populations in high-income countries, may offer an improved approach for surveying hard-to-reach internal migrants in LMIC. What modifications, if any, are required to implement RDS successfully in limited-resource settings? What lessons implementing RDS among migrants in Ghana may be relevant in other LMIC settings? Using rural–urban female internal migrants in Accra, Ghana as a case study, the objectives of this paper are to: 1) assess RDS recruitment productivity, network size, and ties of internal

migrants; 2) test for homophily; and 3) detail the successes of and challenges to implementing RDS in Ghana and how these lessons can be applied in other LMIC settings.

1.1 Surveying migrants with respondent-driven sampling

RDS, a variation on chain-referral sampling methods like snowball sampling, originated in the United States of America (USA) as a method to sample 'hidden' populations such as people living with HIV/AIDS and injecting drug users (Heckathorn 1997). Traditional chain-referral methods may generate biased samples through the oversampling of cooperative participants with larger networks, for example. Such biases would negatively impact inferences from the sample. RDS, however, employs a mathematical model rooted in Markov chain and biased network theories to account for the nonrandom way in which the sample was collected and to generate a weighted sample (Heckathorn 1997). This weighted sample has been proven to be unbiased for samples of meaningful size, regardless of how the researcher purposively selects the initial 'seeds' who initiate the recruitment process (Heckathorn 1997; Salganik and Heckathorn 2004). RDS also relies upon a dual-incentive structure whereby participants are rewarded for both participating in the study and for recruiting additional members of the target population via the distribution of uniquely numbered recruitment coupons. For RDS to succeed, the population of interest must be well networked, with individuals recognizing one another as part of a social group. Additionally, researchers must record each participant's personal network size (the number of eligible participants within each participant's social network).

Use of RDS, and evidence generated from RDS samples, has grown rapidly since the mid-1990s, with over 460 studies in 69 countries (White et al. 2015). The methodology has been revised over the years, leading to adjustments such as the calculation of standard errors, new estimators, a bootstrap method for constructing confidence intervals, and larger design effects (Heckathorn 2002; Salganik 2006; Heckathorn 2007; Volz and Heckathorn 2008; Wejnert et al. 2012). Despite these revisions, the validity of RDS methodology warrants caution and may be inappropriate for certain studies. Simulations find that the RDS assumptions, discussed later in section 2.4, may be unrealistic and can produce biased estimates (Gile and Handcock 2010). Public health researchers in particular may find these estimates and RDS's "misleadingly narrow" confidence intervals unsuitable for disease surveillance (Goel and Salganik 2010: 6746). Biases may also exist around recruitment, causing several researchers to inquire about willingness to participate or ability to recruit (Carballo-Diéguez et al. 2011; Liu et al. 2012; Rudolph, Fuller, and Latkin 2013). A small but

growing body of RDS literature addresses the need to "identify potential recruitment biases so that they can be either acknowledged as potential limitations or corrected for in the analysis" (Rudolph, Fuller, and Latkin 2013: 5).

Researchers have used RDS most frequently in the USA, China, India, Mexico, and South Africa (White et al. 2015). Though four of the five most common countries for RDS studies are classified as LMIC, the use of RDS in migration research is often restricted to sampling hard-to-reach international migrant populations in high-income countries (Tyldum and Johnston 2014). There is a small but growing body of evidence that uses RDS to sample migrants in lower-middle income countries (e.g., Ukraine, Morocco, Kenya) and upper-middle income countries (e.g., South Africa, China, Thailand) (International Organization for Migration 2010; Khamsiriwatchara et al. 2011; Tyldum and Johnston 2014; Okal 2016). In the limited evidence on the use of RDS for migrant populations outside of high-income countries, there is little detailed methodological reflection – either positive or negative – on the suitability of this technique for migrant populations in LMIC settings and the lessons learned.

Researchers have implemented RDS within migrant populations to study issues such as sexual violence among Sub-Saharan migrants in Morocco; HIV risk behaviours among undocumented immigrant women in Texas, USA; and the health of adolescent migrants in Shanghai, China (Montealegre et al. 2013; Decker et al. 2014; Keygnaert et al. 2014). These studies contribute to a nascent body of evidence within the migration literature that documents efforts to ground RDS in local cultures and settings. Keygnaert et al. (2014) adopt a community-based participatory research approach with a community advisory board that contributes to the study design and helps identify community-based researchers in Morocco. Montealegre et al. (2013) incorporate extensive formative evidence in the design of RDS among Central American immigrant women in the USA that includes in-depth interviews with women to assess the acceptability of RDS, to examine social network characteristics, and to pilot the survey questionnaire. Decker et al.'s (2014) five-country (United States, Nigeria, South Africa, India, China) comparative study of disadvantaged 15–19-year-olds also incorporates an extensive formative phase to inform RDS design and implementation, including community mapping, in-depth interviews and focus group discussions with the target population, and key informant interviews.

Some studies using RDS with migrants have reported challenges from slow recruitment, lack of connectedness to the target group, small network size, and reluctance to refer new participants. A study among Pakistani and Polish migrants who recently arrived in the United Kingdom, for example, struggled with chain referral; as a result, researchers developed and implemented an alternative sampling approach in which they led the chain referral process, rather than the participants themselves (Platt, Luthra, and Frere-Smith 2015). This study also found a lack of connectedness of the

target populations as well as participant reluctance to refer additional participants (Platt, Luthra, and Frere-Smith 2015). Among socially marginalized young people in Sydney, Australia, Bryant (2014) found that RDS sampled the target population inefficiently, leading to a high proportion of ineligible participants. Compared to other RDS studies the recruitment rate was low, and participants had small and disparate networks (Bryant 2014).

These documented challenges implementing RDS shed important light on some of the pitfalls involved in using this sampling method for migrant populations. Taken as a whole, they suggest a need to more fully ground RDS in formative research prior to sampling and to share lessons learned from applying RDS in various settings. Involving members of the target population in the design of RDS might help minimize potential implementation challenges such as inappropriate incentives, cultural attitudes towards research participation and peer-recruitment, and whether or not sufficient, direct links exist between members of the target social group. Feedback from the target population and stakeholders may also indicate that RDS may be an inappropriate tool in certain settings or within certain populations. In Moscow, Russia, preliminary fieldwork revealed that irregular female migrants were unable or unwilling to refer participants via RDS; thus, the researchers modified their initial plans and implemented an alternative sampling method (Agadjanian and Zotova 2012).

Little evidence has been published regarding the nuances and practicalities of implementing RDS among migrants in LMIC. With inconsistent documentation of RDS implementation experiences it is difficult to assess where and why RDS is working or failing. The use of RDS in such settings warrants better understanding of factors leading to its successful implementation.

1.2 Sampling female migrants in Ghana

This case study uses data collected during the implementation of RDS among a subpopulation of female internal migrants in Ghana. These migrants are identified by their occupation as market porters who carry loads on their heads (*kayayei*[2] [plural], *kayayoo* [singular]). They represent the face of contemporary female north–south, rural–urban migration in Ghana, with most migrant female youth taking up this work upon arrival in Accra (Kwankye and Addoquaye Tagoe 2009). Yet head porter work is not limited to female migrants from northern Ghana. Kayayei may come from Ghana's

---

[2] *Kayayei* and *kayayoo* are compound words formed from the Hausa and Ga languages. *Kaya* is often translated from Hausa as "load" or "burden." In Ga, *yei* means "women," and *yoo* means "woman." Participants in this study most often used the single Hausa term *kaya* to refer to their work, as in "I carry *kaya*."

other regions and from Togo. Ga petty traders in Accra may even work occasionally as head porters if they cannot make enough money trading (Salifu 2015). In Accra and Kumasi, Ghana's two largest cities, authorities and local residents alike often view kayayei as outsiders, since the majority of kayayei migrated from rural areas in Ghana's northern three regions. Kayayei are highly stigmatized due to a combination of their menial work and northern ethnicities, and they may regard strangers and authority figures with fear and suspicion (Daily Graphic 2010).

These internal migrant kayayei tend to be missed or excluded from household- and institution-based (e.g., school, health facility) surveys since they are mobile, occasionally homeless, and have low levels of formal education. Though vulnerable to sexual and reproductive health risks and violence, migrant kayayei seek formal health services infrequently (Anarfi and Appiah 2009; Kwankye and Addoquaye Tagoe 2009), excluding the possibility of facility-based sampling methods to research health-related issues. This population has become a growing concern for the government, nongovernmental organizations, and civil society organizations. The growing migration of girls and women from northern Ghana seeking kayayei work to avoid poverty highlights a need for greater development in the north. Large numbers of teenagers working as kayayei challenge Ghana's efforts to combat child labour. The increasing number of adolescents and young adults in Accra places additional pressure on social services and employment opportunities. Despite consensus that the total number of kayayei in Accra is increasing, the exact figure remains unknown: estimates range from 2,300 to 160,000 (Kearney 2013; Parliament of the Republic of Ghana 2016).

Researchers have identified multiple challenges gathering data on Ghana's kayayei, including a lack of incentives to participate, language barriers, and sampling and survey approaches. Without compensating participants for their time, Opare (2003) found kayayei unlikely to participate in research lasting over five minutes. In the country's 2010 Population and Housing Census the number of kayayei was higher than anticipated and required more than one day to enumerate (Daily Express 2010). According to media reports and formative interviews with kayayei leaders, census implementation efforts were further challenged by attempts to enumerate kayayei whilst they were working and by waking up sleeping homeless kayayei in an effort to enumerate them (Daily Graphic 2010). Given the stigma that these migrants experience in Accra and political discussions about forcibly sending migrant kayayei back north, many migrant kayayei hesitate to interact with unknown people such as census enumerators, who represent authority. Convenience sampling of kayayei in the market workplace, for example, would cause an unwelcome interruption to participants' work, especially during busy market times when research participation could result in the loss of potential customers. In order to have large networks in Accra, migrant kayayei would likely have resided in Accra for a long time or have migrated multiple times; therefore,

with chain-referral methods like snowball sampling, recruiting migrant kayayei with the largest networks would likely lead to bias.

Despite the methodological and implementation limitations of RDS, this method offers an opportunity to overcome previously documented kayayei sampling challenges (e.g., declining to participate when interrupted at work, lack of a sampling frame) and an opportunity to generate stronger population-level inferences than prior studies employing nonrandom sampling methods. By design, RDS features (e.g., peer recruiting, coupons) could also help facilitate participant willingness and availability, which have hindered prior efforts to sampling migrant kayayei. Ghana's migrant kayayei meet the criteria to sample using RDS. Migrant kayayei can be clearly defined, recognize one another as part of the group, and have the characteristics of a social group (i.e., they identify and interact with each other) (Friberg and Horst 2014). Formative discussions with key informants, including members of the Kayayei Youth Association (KYA) and kayayei leaders, indicated that kayayei migrants would have sufficiently dense network ties to facilitate chain referral. Organizations in Accra offer events and programmes (e.g., community health outreach programmes, skills training, activities organized by the KYA) to kayayei regardless of ethnolinguistic group, providing opportunities for kayayei to meet and interact with members of different tribes. North–south migration in Ghana is a highly network-driven phenomenon with migrant kayayei linked by kinship, friendship, community, and tribe. Kayayei are directly linked to one another, and these connections impact their migration. For girls and women in the north considering migrating to Accra, relationships with current migrants influences choice of destination. Kayayei migrants work together in groups, sleep together outside or in a shared room, watch over each other's children, and support one another.

2. Data and methods

Quantitative data was collected using RDS over four weeks in March and April 2015 as part of a larger mixed-methods study on migration and women's health in Ghana. Given this larger study's focus on north–south migration, the quantitative survey was limited to kayayei in Accra who migrated from Ghana's three northern regions. The study was conducted at the KYA office in Agbogbloshie, Accra, near several of the capital's largest markets and Accra's largest informal settlement. Based on previous researchers' experiences attempting to interview kayayei, the choice of research location was identified a priori as critical to the success of the RDS. Situating the research out of the KYA office, where the research team was visible to anyone walking by, also served to allay the fears of potential recruits, who could observe the research team's activities from a comfortable distance before deciding to participate. In addition, working out of

the KYA's office allowed the research team to link participants needing additional support with services that the KYA provides in collaboration with partners (e.g., Marie Stopes Ghana, United States Agency for International Development, Society for Women and AIDS in Africa). In recognition of the diverse roles that female migrants play, including that of mother and caregiver, the office was a child-friendly place. Infants and young children accompanied participants on a daily basis.

2.1 Formative research

Before implementing the survey, the author spent two months in the community conducting formative research and volunteering with the KYA. Formative activities identified as critical prior to implementing RDS included meetings with a wide range of stakeholders (e.g., kayayei community leaders, potential seeds who would initiate the study, KYA board members) to introduce the study, seek their input, and answer questions about the research. The formative research generated evidence on factors necessary for implementation of RDS, including appropriate incentives; participants' personal network sizes; hours and days during which to conduct the study; coupon design; the survey instrument; participants' languages; and participants' prior experiences with researchers, outsiders, and census enumerators. People were unfamiliar yet intrigued by the idea of using coupons to recruit and to participate. Formative discussions about the use of coupons revealed that, so long as the incentives were appropriate, kayayei migrants would participate in longer interviews. Provided participants received an appropriate incentive that compensated them for missed work and travel, key informants expected that participants would sit for at least 45 minutes to complete the survey.

Since the KYA office was open to participants on Mondays to Saturdays from 9:00am to 5:00pm, when many kayayei migrants were working, participants received monetary incentives relative to a kayayoo's average daily wage. The office's location near large markets made it feasible for kayayei to participate in research during breaks and at times when slow markets produced little demand for their services. Stakeholders, including the KYA and working kayayei, assisted with identifying the most appropriate incentive values during the formative research period. Incentives, a key feature of the RDS methodology, functioned as both economic compensation for participation and motivation for participants to recruit additional respondents (Tyldum et al. 2014). Survey respondents received a total of 5 Ghanaian cedis (approximately £0.85 as of 13 November 2017): 2 cedis for participation and 3 cedis for travel and transportation. Respondents were eligible to receive a bonus of up to 3 cedis (approximately £0.51 as of 13 November 2017) for referring up to three additional participants.

To accurately measure personal network size (PNS) – the number of eligible participants within each participant's social network – the formative assessment phase explored what it meant to know someone in this setting, and piloted ways in which to ask participants these questions. Capturing each participant's PNS is crucial for analysing data using RDS and requires: 1) a clear definition of the target population, 2) a defined meaning of what it means to know someone, 3) a geographic parameter, and 4) a time frame (Johnston, Rodriguez, and Napierala 2014).

The critical importance of the identity of interviewers also emerged during the formative work. Key informants recommended hiring local data collectors who identified as insiders rather than outsiders and could thus minimize social distance (Driedger and Peters 1977) and put the study population at greater ease. Additionally, kayayei data collectors would be well positioned to confirm eligibility criteria and rule out potential imposters, such as beggars, who might attempt to participate for the incentive. Seven data collectors from northern Ghana who self-identified as kayayei and who had a high school education were selected and trained to administer the survey questionnaire to participants. Data collectors spoke English, Twi, Dagbani, Frafra, Mampruli, Kusaal, and Sisaala. Like the kayayei community leaders consulted during the formative research phase, all data collectors were bilingual or multilingual, speaking two (n = 2), four (n = 2), or five languages (n = 3). The number of data collectors was based on linguistic needs as well as sampling estimates that assisted in study planning and implementation (Wejnert et al. 2012). Conservative calculations revealed that an estimated sample of 538 kayayei was required to achieve the same statistical power as a simple random sample one-tenth as large (Lattof 2018).

2.2 Recruitment of seeds and participants

Following this formative work and a review of studies with comparable estimated sample sizes, ten seeds were initially identified in collaboration with the KYA. Whilst no exact method exists for selecting seeds, prior studies have often selected 1–20 seeds that reflect diversity (Kubal, Shvab, and Wojtynska 2014). The author's analyses of census data as well as prior studies on kayayei indicate that most north–south female kayayei migrants are Muslims from the Northern Region, so a higher proportion of seeds reflected these characteristics. Seeds were selected to ensure variety across additional characteristics, such as the current length of time spent in Accra, the number of migrations made to Accra, primary worksite in Accra, and homelessness. Four seeds failed to generate sufficient recruitment chains, producing an insufficient number of referrals to generate a final sample independent of these four seeds. Two of those failed seeds identified as Mole-Dagbani, an ethnic group that was well represented among the

remaining seeds. The other two failed seeds identified as Sissala and Frafra, warranting replacement seeds that identified similarly to ensure sufficient representation. The remaining six seeds were growing sufficiently and their recruitment chains continued to generate new recruits, so only two replacement seeds were added to the study.

In consultation with stakeholders, coupons were designed and printed in English, the primary language of instruction in Ghana's schools. Given the low levels of education and literacy among the study population, many participants would likely be unable to read the coupons. If illiterate participants sought assistance from someone outside the office, whoever read the coupon would have learned English in school. To minimize issues with using coupons among a largely illiterate population for whom the coupons would serve as a visible token of potential participation, data collector training emphasized the importance of concise oral recruitment instructions to each participant after the survey. Participants were encouraged to return to the office if they had questions about the coupon details. Participants received three recruitment coupons until their recruitment chain reached the third wave, at which point the number of coupons was reduced to two. In the last week of the study the number of recruiting coupons was gradually reduced to zero. Initially, coupons expired within one week in order to minimize coupon degradation and maximize the possibility that recruiters would correctly recall the recruitment instructions. It became apparent after the first week that a shorter referral window would lead to less coupon deterioration and fewer reports of coupons being taken north with potential recruits returning home. The coupon validity period was reduced to three days for the remainder of the survey.

Recruits who held a valid coupon were invited to participate in the study if they met the following selection criteria:

1. Female
2. Currently working as a kayayoo in Accra
3. Migrated to Accra from one of the three northern regions (Upper East Region, Upper West Region, or Northern Region)
4. Has not already participated.

Assigning participants to data collectors based on shared ethnic group and/or shared language increased the likelihood that any ineligible returning participants would be reinterviewed and identified by their original data collector. Whilst data collectors were advised to be aware of the possibility of repeat participants and were encouraged to report suspected repeat participants, they identified no cases.

The data collection team ensured that prior to participation all participants received the required study information in a manner that was understandable, so that consent was obtained without coercion or undue influence. Due to low literacy levels among the

study population and suspicion over signing documents, verbal consent was preferable to written consent so as to reduce potential embarrassment and underscore the participants' agency. Participants were offered the option of verbal consent in which the researchers followed a set script (Evans and Becker 2009; GSS, GHS, and ICF Macro 2009; Day 2014). Among participants under 18 years of age, the data collectors inquired whether the participants migrated to Accra with a parent or guardian or if they were 'emancipated minors' who migrated to Accra alone. Adults (those aged 18 years and older) and independent child migrants (emancipated minors under 18 years of age) received the standard consent form. Children under 18 years who resided in Accra with a parent or guardian were asked to bring their parent or guardian to the office to complete a parental consent form and child assent form before participating. The London School of Economics (LSE) Research Ethics Committee (on 18 November 2014) and the University of Ghana's Noguchi Memorial Institute Institutional Review Board (protocol number 065/14-15) approved this research. In addition, the author sought approval from kayayei community leaders, the KYA's Board of Directors, the Director of Social Welfare, and the local chief.

2.3 Data collection

The research team pretested and piloted the questionnaire during the formative research period in order to verify the questionnaire structure, formulation of questions, appropriate language and terminology, completion times, range of variation in response variables, and respondent understanding. Data collectors found that the term 'sister' was causing confusion, since in this setting 'sister' is often used to refer to a close female friend. All survey references to 'sister' were thus changed to 'biological sister.' Data was recorded in a 25-page questionnaire with ten modules (personal network size, demographic characteristics, employment, reproduction, recent illness or injury, health insurance, relationships, family planning, sexually transmitted infections, and data collector observations). Though the majority of training concluded before the study started, the author continued on-the-job training during survey implementation.

Since RDS analysis relies on the collection of the personal network size (PNS) variable that measures the number of eligible people in each participant's personal network who are members of the study's target population (Johnston, Rodriguez, and Napierala 2014), data collector training emphasized what it means to know someone and different ways of phrasing this expression, such as "someone who would recognize you at the market." These relationships between recruiters and participants should include strong and weak ties in order to sample a wide range of participants from the target population and reduce the likelihood of recruitment bottlenecks (Johnston,

Rodriguez, and Napierala 2014). Data collectors obtained participants' PNS from three questions at the start of the survey. The first question defines the target population and what it means to know someone (#101). In order to prevent the recruitment of strangers, this question also references time spent together (e.g., carrying loads together for the same customer, sharing a room, sitting together on a break). Subsequent PNS questions limit the survey's geographic boundary to Accra (#102) and the estimation period to two weeks (#103) to reduce recall bias.

101. How many kayayei do you know who are from the northern three regions: Upper East, Upper West, and Northern Region? By 'know,' I mean you know them and they know you. You know their name, and they know your name. They are people you have spent time with who would recognize you in the market.

102. How many of these girls and women currently work in Accra as kayayei? Please give an exact number.

103. How many of these northern kayayei girls and women have you seen at least once in Accra during the last two weeks? Please take your time to count so that you can give me an exact number.

The answer to this third question is considered to be the participant's PNS and is used in analyses to weight the data and assess sample bias. It is from this network that respondents ideally recruit at random; however, kayayei migrants reported during the formative assessment phase that their networks were highly mobile, sceptical or fearful of participating in research, and might believe the coupons were a scam. RDS permits data capture on nonrespondents when participants return to collect their recruiting bonuses, shedding light on specific factors that affect a population's willingness to participate. Based on this information and evidence in the literature on recruitment biases (Liu et al. 2012; Rudolph, Fuller, Latkin 2013), a fourth PNS question was added to the survey questionnaire in an attempt to acknowledge participants' inherent existing biases and more accurately measure PNS. When compared to PNS question 103, this fourth PNS question may help identify or minimize the potential effects of participants not recruiting individuals within their networks randomly (e.g., not recruiting individuals who plan to migrate soon, recruiting someone who needs money, not discussing the research with sceptical individuals who might decline a coupon). The fourth question emphasizes which contacts in a respondent's network she would consider inviting to participate in this research:

104. From the number you just told me, how many of these kayayei would you consider inviting to this study?

Quality checks were built into the survey and participant flow. Once the data collector finished administering a survey, the author reviewed the survey on site for completeness and accuracy and resolved any inconsistencies or skipped questions with the participant before she received her payment and recruitment instructions. A subset of purposively selected survey participants (n = 48) completed in-depth interviews to discuss their migration histories and their experiences with illness/injury in the two weeks preceding the survey. The author conducted these interviews with a translator who was identified from amongst the data collectors when her skill at connecting with participants became apparent during the survey. The interviews, recorded and transcribed in English, also serve as an important indirect source of information on participants' reasons for participating in social science research.

2.4 Analyses

Analyses of the population parameters, including the final sample size and sample stability, were conducted using Respondent Driven Sampling Analysis Tool (RDSAT) Version 7.1 (Volz et al. 2012). Apart from analyses of recruitment productivity, seeds were excluded from additional analyses. Partition analysis for categorical and continuous variables in RDSAT used the following recommended options for optimal accuracy and the most stable estimates: dual-component estimate for average network size with a mean cell size of 12, enhanced data-smoothing algorithm type, 16,000 resamples for Bootstrap, and alpha = 0.025 (Heckathorn 2007; Spiller, Cameron, and Heckathorn 2012). Prevalence estimation and equilibrium, or the point at which the sample distribution within variables has stabilized and results in a sample independent of the nonrandom kayayei migrant seeds (Heckathorn 1997, 2002), were also assessed in RDSAT. As recommended in the RDSAT 7.1 User Manual, estimates of the number of waves required for equilibrium used the default convergence radius of 0.02 (Spiller, Cameron, and Heckathorn 2012). Results are reported based on the criteria in the Strengthening the Reporting of Observational Studies in Epidemiology for RDS Studies (STROBE-RDS) checklist (White et al. 2015).

These analyses are based on six assumptions about RDS: 1) respondents have reciprocal relationships with one another; 2) respondents' social networks are dense enough to sustain a chain-referral process; 3) each respondent recruits a single peer; 4) respondents recruit randomly from their networks; 5) respondents can accurately report their personal network size to data collectors; and 6) sampling occurs with replacement

(Salganik and Heckathorn 2004; Heckathorn 2007). Pre-implementation discussions with kayayei generated confidence in the first three assumptions. The fourth assumption (respondents recruit randomly from their networks) is difficult to ensure since factors such as the incentives, study location(s), ethnolinguistic identity, and research topic may influence to whom participants give their coupons (Heckathorn 2007). The researcher discussed these factors with key informants during the formative research period when assessing the appropriateness of RDS and selecting both the study location and incentives. Whilst this fourth assumption is a limitation of RDS that cannot currently be rectified, the survey questionnaire included a fourth question on participants' personal network sizes in an attempt to minimize potential violation of the fourth assumption, specifically potential recruitment biases underlying the recruiter–recruit relationship. This question and the fifth assumption (accurately reporting personal network size) were tested extensively during data collector training. Most RDS studies violate the sixth assumption (sampling occurs with replacement) since, by design, participants cannot participate again. Biases from violating this particular assumption remain poorly understood (Volz and Heckathorn 2008; Goel and Salganik 2009). Among samples that are relatively small compared to the total population's size, the sixth assumption is relaxed (Volz and Heckathorn 2008).

3. Results

Of 643 potential participants recruited using RDS, including 12 seeds, a total of 625 participants completed the survey (Figure 1). Participants represented fifteen different ethnic groups and had a mean age of 25.2 years (s.d. 8.1 years). One in ten participants completed the survey in Twi, a southern dialect of Akan, and 4.0% of participants completed the survey in English. In-depth interview participants reported learning Twi as quickly as possible upon arrival in Accra in order to communicate with customers and obtain work.

Figure 1: **RDS participant flow diagram of kayayei migrants in Accra, Ghana (March to April 2015)**
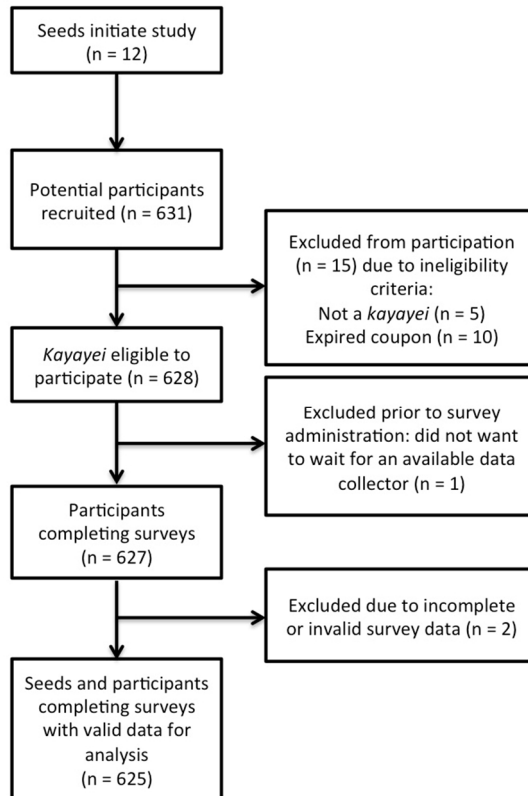
Nearly all participants (97.6%) cited employment as a primary reason for migrating to Accra, and 93.0% of participants knew that they would work as kayayei before migrating. Over half of participants (56.0%) migrated to Accra alone. Participants reported migrating to Accra between one and seven times. One-third of participants were on their first migration and one-third of participants were on their second migration. The remaining third of participants migrated to Accra three or more times. Of the 544/625 participants who recommended kayayei work to someone in the north, 541 participants (99.6%) reported that the individuals to whom they recommended kayayei work later migrated south. Only 3.2% of participants had prior

kayayei experience outside of Accra: three participants previously worked as kayayei in Kumasi (Ashanti Region), and 17 participants previously worked as kayayei in Tamale (Northern Region).

3.1 Recruitment productivity

An average of 156 recruits completed the survey each week. Excluding the 12 seeds recruited into the study without coupons, participants redeemed 613 of the 915 recruitment coupons (67.0%) distributed. This coupon return rate exceeds that of other studies using RDS among migrants; these studies generally report coupon redemption rates from 30.0% to 45.0% among migrant populations in Australia, Albania, and Kenya (Johnston et al. 2010; International Organization for Migration 2010; Bryant 2014). The high rate of return achieved in this study is more comparable to studies examining adolescent health among nonmigrant populations in Ibadan, Nigeria (53.7%) and Delhi, India (75.0%) (Decker et al. 2014).

Seeds two, three, and eleven generated the largest numbers of linked recruits (Table 1). The longest recruitment chain consisted of 11 waves. Four of the ten initial seeds failed to generate recruitment trees. The youngest seeds (#6 and #7) (aged 12 and 13 years, respectively) recruited no one. Age may be a critical factor when selecting seeds, though the most productive seed (#2) was 15 years old. Seed nine recruited only one participant. These three unproductive seeds (#6, #7, and #9) failed to recruit for reasons that remain unknown. Seed eight recruited two participants before that recruitment tree failed. After that tree stopped generating new recruits, the seed informed us that one of her recruits who accepted a coupon had returned to the Upper East with the coupon. Seed eight mentioned that her other two successful recruits had given out all their coupons, but then people in the community told these potential recruits that they were being fooled and should not participate. It is possible that the younger seeds also received such 'advice' from older community members who felt they were acting in the seeds' best interests. In response to these failed seeds, two older replacement seeds (#11 and #12) were recruited. At the end of the study, eight of the twelve seeds generated recruitment chains of sufficient length for analysis.

Table 1: **Characteristics of seeds sorted by increasing recruitment productivity**

	Seed characteristics										Seed recruitment productivity		
#	Original or replacement seed	Age in years	Ethnic group	Region of origin	Number of migrations to Accra	Current length of time in Accra	Worksite in Accra	Homeless	Highest level of school completed	Religion	Number of seed's direct recruits	Number of waves linked	Total number of recruits linked (%)
6	O	12	Mole-Dagbani	Northern	1	2 months	Tema Station	No	Primary	Muslim	0	0	0 (0.0)
7	O	13	Mole-Dagbani	Northern	1	1 year	Makola Market	No	None	Muslim	0	0	0 (0.0)
9	O	27	Sissala	Upper West	1	1 year	Plantain Market	No	None	Muslim	1	1	1 (0.16)
8	O	20	Frafra	Upper East	2	7 months	Agbogbloshie	No	Middle	Catholic	2	1	2 (0.33)
12	R	27	Frafra	Upper East	4	3 years	Makola Market	No	None	Presbyterian	3	11	45 (7.34)
5	O	28	Mole-Dagbani	Northern	3	8 months	Roams around	No	None	Muslim	3	9	51 (8.32)
1	O	32	Mampruli	Northern	1	7 years	Makola Market	Yes	None	Muslim	3	9	53 (8.65)
4	O	14	Mole-Dagbani	Northern	1	1 year	C.M.B.	Yes	Primary	Muslim	3	10	58 (9.46)
10	O	19	Kussasi	Upper East	2	4 months	Darkuman	No	Primary	Muslim	3	8	61 (9.96)
11	R	30	Sissala	Upper West	4	7 years	Agbogbloshie	No	None	Muslim	2	10	97 (15.82)
3	O	25	Mole-Dagbani	Northern	5	8 months	Konkomba Market	Yes	None	Muslim	3	10	117 (19.09)
2	O	15	Mole-Dagbani	Northern	1	4 months	Konkomba Market	No	Primary	Muslim	3	10	136 (22.19)

The recruiter–recruit relationship was tracked with a series of nine questions that were incorporated into the survey. Participants most often reported that their recruiters were friends (58.6%) or biological sisters (25.8%). The majority of participants (95.0%) saw their recruiters every day and knew them very well (96.3%), indicating strong preexisting networks. Three in five participants had known their recruiter for one or more years. Two in five participants knew their recruiter for ten months or less. The survey also asked recruits what type of work their recruiters did for further verification that the recruiters were indeed kayayei migrants.

Data collectors asked participants returning to collect recruitment payments if anyone declined to accept a participation coupon. Participants provided nonresponse data for 16 potential recruits who ranged in age from 16 to 40 years. Participants had strong relationships with these potential recruits and saw them nearly every day. About one third of nonrespondents declined a participation coupon on the grounds that they were too busy to participate. Fear of research was the second most common reason (25.0%) for declining a coupon, followed by lack of interest (18.8%).

3.2 Network size and ties

To compare the effect of different questions capturing personal network size, this section reports outcomes calculated from both the standard PNS question (#103) and the fourth PNS question (#104) added in an attempt to provide a potentially more robust measure in this setting. Participants' mean network size of 22.73 kayayei (s.d. 13.55, min. 1, max. 99) decreased to 9.66 kayayei (s.d. 6.68, min. 1, max. 50) when asked to specify how many individuals in their network they would consider inviting to this study (Table 2). Network size increases as participants become more educated and as the number of migrations to Accra increases (Table 3). Homeless participants reported larger networks than participants who were not homeless.

Table 2:   Comparison of PNS questions and participants' mean network sizes

PNS question	Question's aim	Participants' mean network size (s.d.)
101. How many kayayei do you know who are from the northern three regions: Upper East, Upper West, and Northern Region? By 'know,' I mean you know them and they know you. You know their name, and they know your name. They are people you have spent time with who would recognize you in the market.	Define the target population and what it means to know someone	88.03 (51.43)
102. How many of these girls and women currently work in Accra as kayayei? Please give an exact number.	Limit the survey's geographic boundary to Accra	49.04 (26.50)
103. How many of these northern kayayei girls and women have you seen at least once in Accra during the last two weeks? Please take your time to count so that you can give me an exact number.	Limit the estimation period to two weeks to reduce recall bias	22.73 (13.55)
104. From the number you just told me, how many of these kayayei would you consider inviting to this study?	Help identify or minimize the potential effects of participants not recruiting individuals within their networks randomly	9.67 (6.68)

Ethnic group and religion, two variables that did not obtain equilibrium, appear to have restricted some respondents from participating. Ethnic ties remain strong for many participants. Homophily indices (Table 3), a measure of cross-group recruitment ranging from –1 (preference for recruiting from outside one's group) to +1 (preference for recruiting from within one's group) (Heckathorn 2002), show that participants belonging to Mampruli, Sisaala, Mole-Dagbani, and Frafra tribes were unlikely to recruit across ethnic group; these participants strongly preferred to recruit members of their own tribes. This unwillingness to recruit across ethnic group may be rooted in linguistic differences. It could also reflect a desire to share resources (i.e., the RDS incentives) internally with close friends and family. Homophily indices below 0.4 indicate cross-group recruitment, whereby participants with a particular characteristic are willing to recruit participants without that particular characteristic. Participants recruited freely across characteristics like education level, homelessness, age, and recent illness/injury.

To understand how changes in capturing PNS affect the analyses, the author calculated the prevalence of recent illness and/or injury. The prevalence of illness and/or injury in the past two weeks among kayayei migrants in Accra is 27.4% (confidence interval 23.6% to 33.1%) using the PNS data captured in the standard (three-part) PNS question. This figure declines to 23.1% (confidence interval 19.6% to 28.1%) when analysed using the more conservative PNS data incorporating the fourth question.

Table 3: **Effect of personal network size questions on network sizes and homophily for key demographic characteristics**

Characteristic	Q103 personal network size		Q104 personal network size	
	Adjusted average network size	Homophily index	Adjusted average network size	Homophily index
Education level				
None	12.147	−0.015	5.065	0.015
Primary	16.230	0.208	6.932	0.214
Middle/JSS	19.103	0.066	5.347	0.042
Secondary/SSS	22.133	0.132	11.130	0.145
Age group (in years)				
10–14	11.502	0.176	4.189	0.161
15–19	10.762	0.057	5.123	0.087
20–24	14.001	0.097	5.465	0.085
25–29	13.570	0.007	5.464	0.002
30–34	12.388	0.084	5.096	0.077
35–39	18.920	0.047	8.432	0.076
40–44	18.920	0.026	6.560	0.018
45–49	16.347	−1.000	8.737	−1.000
50+	9.019	−1.000	3.540	−1.000
Homeless				
Yes	19.441	0.169	6.463	0.144
No	12.349	−0.042	5.212	−0.013
Current migration to Accra				
1st	12.734	0.057	5.223	0.063
2nd	12.585	0.022	5.450	0.057
3rd	12.961	0.011	5.100	0.002
4th	17.036	0.041	5.208	0.024
5th	17.450	−1.000	8.055	−1.000
6th	16.984	−1.000	6.781	−1.000
7th	18.837	−1.000	8.291	−1.000
Religion				
Catholic	22.911	0.566	10.011	0.567
Anglican	18.374	−0.999	11.201	−0.999
Methodist	19.773	−0.995	12.593	−0.991
Presbyterian	28.160	0.067	11.622	0.067
Pentecostal/Charismatic	24.977	−0.999	9.968	−0.999
Muslim	12.502	0.489	5.093	0.461
Traditional/Spiritualist	7.000	0.140	2.000	0.138
Ethnic group				
Mampruli	17.208	0.739	7.486	0.740
Sissala	16.232	0.917	7.284	0.917
Ashanti	20.000	0.066	15.000	0.066
Guan	20.000	−0.997	8.000	−0.997
Mole-Dagbani	11.530	0.766	4.620	0.747
Grussi	24.262	0.179	9.698	0.179
Gruma	25.714	−0.987	8.699	−0.989
Konkomba	6.173	−1.000	4.432	−1.000
Dagaba	40.000	0.066	29.000	0.066
Kusasi	21.951	0.165	10.313	0.166
Mandé	55.000	0.066	7.000	0.066
Frafra	27.552	0.724	11.928	0.724
Walla	30.000	−0.970	20.000	−0.950
Gonja	30.000	−0.990	9.000	−0.992
Hausa-Dagomba	11.000	0.064	2.000	0.061

Table 3: **(Continued)**

Characteristic	Q103 personal network size		Q104 personal network size	
	Adjusted average network size	Homophily index	Adjusted average network size	Homophily index
Illness or injury in last two weeks	18.034	0.207	8.908	0.254
Illness	17.915	0.028	6.231	0.026
Injury	28.713	−0.999	11.591	−0.999
Both	11.545	−0.084	4.439	−0.134
Neither (healthy)				

In addition to changes in prevalence, differences in capturing PNS result in subtle changes to estimated population proportions (Table 4). For age group (in years), asking a narrower PNS question (#104) reduced the waves needed for equilibrium from five to four, meaning that after four waves the sample distribution within age group stabilized and produced a sample independent of the age groups of the initial nonrandom migrant kayayei seeds (Heckathorn 1997, 2002). Upon further inspection of the data, a decline in waves required for equilibrium also appeared for respondents' primary source of drinking water; however, an opposite trend appeared for respondents' relationship to their recruiter and place of residence in Accra. Both variables required one additional wave to reach equilibrium when estimates used question #104. Asking PNS question (#104) in an attempt to exclude any known or perceived nonresponders from the recruiter's network calculation certainly did not harm the RDS population estimates.

The majority of variables achieved equilibrium in five waves or fewer. Ethnic group, however, was a notable exception. Achieving equilibrium for this variable would have required running the survey for an average of 54 waves (min 48, max 80), exceeding the resources of this study. It was also not possible to achieve equilibrium for religion or place of residence in Accra, two variables that are closely affiliated with ethnic group in this setting.

Table 4: **Sample proportions for demographic characteristics versus population estimates by personal network size question**

Characteristic	Sample population proportions	Q103 personal network size			Q104 personal network size		
		Estimated population proportions (95% CI)	Standard error	Waves required	Estimated population percentages (95% CI)	Standard error	Waves required
Education level				3			3
None	73.3	79.8 (74.8, 84.1)	0.024		78.3 (73.1, 82.7)	0.025	
Primary	17.3	13.8 (10.0, 18.3)	0.021		13.2 (9.8, 17.2)	0.019	
Middle/JSS	7.4	5.1 (3.4, 7.0)	0.009		7.4 (4.6, 10.9)	0.016	
Secondary/SSS	2.1	1.3 (0.5, 2.3)	0.005		1.0 (0.4, 1.8)	0.004	
Age group (in years)				5			4
10–14	6.9	8.0 (4.8, 12.6)	0.020		8.9 (5.4, 13.0)	0.020	
15–19	15.0	18.9 (13.6, 25.0)	0.029		16.2 (12.2, 21.3)	0.023	
20–24	25.9	24.9 (19.3, 29.0)	0.024		26.1 (20.9, 30.6)	0.024	
25–29	19.5	18.7 (14.5, 23.8)	0.024		19.0 (14.8, 23.9)	0.023	
30–34	18.9	19.6 (14.5, 25.2)	0.027		19.6 (15.3, 25.1)	0.025	
35–39	5.6	3.8 (2.2, 5.8)	0.009		3.4 (2.2, 5.2)	0.008	
40–44	6.2	4.2 (2.7, 6.2)	0.009		4.9 (2.8, 6.0)	0.008	
45–49	1.1	0.8 (0.2, 1.9)	0.005		0.6 (0.1, 1.3)	0.003	
50+	0.8	1.1 (0.1, 2.6)	0.007		1.2 (0.1, 2.6)	0.006	
Homeless				2			2
Yes	17.6	11.9 (9.0, 15.1)	0.016		14.4 (11.4, 18.4)	0.018	
No	82.4	88.1 (84.9, 91.0)	0.016		85.6 (81.6, 88.6)	0.018	
Current migration to				2			2
Accra	33.4	34.8 (28.6, 40.8)	0.031		34.8 (29.4, 40.6)	0.028	
1st	35.4	37.2 (31.0, 42.9)	0.030		35.1 (29.0, 39.7)	0.027	
2nd	17.1	17.4 (13.5, 22.8)	0.024		18.2 (14.2, 23.5)	0.024	
3rd	7.0	5.2 (3.4, 7.7)	0.011		6.9 (4.4, 10.1)	0.015	
4th	3.4	2.4 (1.1, 4.3)	0.008		2.1 (0.1, 3.5)	0.006	
5th	2.9	2.4 (1.1, 4.0)	0.007		2.4 (1.3, 3.9)	0.007	
6th	0.8	0.5 (0.1, 1.2)	0.003		0.5 (0.1, 1.2)	0.003	
7th							
Religion				17			17
Catholic	8.6	3.6 (1.6, 7.1)	0.014		3.4 (1.5, 6.6)	0.013	
Anglican	0.5	0.3 (0.0, 0.8)	0.002		0.2 (0.0, 0.5)	0.001	
Methodist	0.3	0.2 (0.0, 0.6)	0.002		0.1 (0.0, 0.4)	0.001	
Presbyterian	1.4	0.5 (0.1, 1.1)	0.003		0.5 (0.1, 1.1)	0.003	
Pentecostal/Charismatic	0.8	0.3 (0.0, 0.7)	0.002		0.3 (0.0, 0.7)	0.002	
Muslim	88.2	94.8 (90.5, 97.5)	0.019		95.0 (91.0, 97.6)	0.021	
Traditional/Spiritualist	0.2	0.4 (0.0, 1.2)	0.008		0.5 (0.0, 1.7)	0.014	

Table 4: **(Continued)**

Characteristic	Sample population proportions	Q103 personal network size			Q104 personal network size		
		Estimated population proportions (95% CI)	Standard error	Waves required	Estimated population percentages (95% CI)	Standard error	Waves required
Ethnic group				56			56
Mampruli	5.9	4.3 (1.1, 8.6)	0.020		4.0 (1.0, 8.0)	0.018	
Sissala	13.3	7.6 (0.5, 20.2)	0.058		6.9 (0.5, 18.7)	0.057	
Ashanti	0.2	0.1 (0.0, 0.5)	0.001		0.1 (0.0, 0.4)	0.004	
Guan	0.2	0.1 (0.0, 0.4)	0.001		0.1 (0.0, 0.4)	0.001	
Mole-Dagbani	67.2	84.7 (69.8, 93.8)	0.075		85.9 (71.7, 94.4)	0.073	
Grussi	1.8	0.3 (0.0, 1.1)	0.010		0.4 (0.0, 1.2)	0.008	
Gruma	0.5	0.1 (0.0, 0.3)	0.005		0.1 (0.0, 0.4)	0.005	
Konkomba	0.5	0.9 (0.0, 3.0)	0.008		0.5 (0.0, 1.8)	0.005	
Dagaba	0.2	0.0 (0.0, 0.4)	0.010		0.0 (0.0, 0.4)	0.003	
Kusasi	0.6	0.2 (0.0, 0.4)	0.012		0.1 (0.0, 0.5)	0.011	
Mandé	0.2	0.0 (0.0, 0.3)	0.001		0.1 (0.0, 0.4)	0.001	
Frafra	9.1	1.2 (0.0, 3.9)	0.035		1.1 (0.0, 3.7)	0.034	
Walla	0.2	0.0 (0.0, 0.3)	0.007		0.0 (0.0, 0.3)	0.001	
Gonja	0.2	0.1 (0.0, 0.3)	0.001		0.1 (0.0, 0.4)	0.001	
Hausa-Dagomba	0.2	0.3 (0.0, 0.9)	0.008		0.6 (0.0, 1.7)	0.009	
Illness or injury in last two weeks				2			2
Illness	35.2	25.9 (22.1, 31.3)	0.024		21.3 (18.0, 26.1)	0.021	
Injury	2.6	1.4 (0.5, 2.7)	0.006		1.6 (0.6, 3.2)	0.007	
Both	0.6	0.1 (0.0, 0.4)	0.001		0.1 (0.0, 0.4)	0.001	
Neither (healthy)	61.6	72.6 (66.9, 76.4)	0.024		76.9 (71.9, 80.4)	0.022	

* Population proportions are reported per 100 individuals.

3.3 Successes and challenges implementing RDS

The study population perceived RDS as a feasible, well-received sampling method, based on feedback from participants and key informants. Primary advantages of this method include: empowering potential participants to participate, if interested, at a time of their choosing; employing data collectors from the study population who participants trusted; and recognising participants' missed work opportunities through participation incentives. Participant clustering and rapid coupon deterioration may limit successful RDS implementation in similar settings. Exploring these factors during a formative research phase with support from community members will ultimately assist in successful implementation.

3.3.1 Coupons and recruitment

Key informants and participants perceived RDS's peer recruitment and coupon features as a means to give participants control over whether or not to participate and the timing of their participation. Participants could complete the survey when it suited their schedules, such as periods when the markets were slow or on breaks. During the first week of data collection, data collectors noted occasional stress caused by participant clustering. Participants often waited to redeem their coupons in groups (e.g., those who share a room, friends who work at the same market) around lunchtime and in the afternoon.

> **Lesson learned:** To manage participant flow, suggest that participants encourage their recruits to come at times when the office is known to be quieter, with shorter waiting times. Offer water and snacks to participants waiting to be surveyed, and provide approximate wait times to participants during busy hours.

The use of paper coupons was difficult in Agbogbloshie, where paper deteriorates quickly. Multiple participants and potential participants reported coupons being burned in building fires. Some potential participants reported that their coupons, stored with their money, had been stolen when thieves robbed them at night. The rainy season also soaked multiple coupons until the participant numbers were nearly illegible.

> **Lesson learned:** Laminating coupons or choosing a thicker paper with indelible ink could help mitigate coupon deterioration under similar circumstances. Shortening expiration dates to two or three days can also help mitigate coupon deterioration to a degree; however, some coupon deterioration and loss is unavoidable in settings with similar climates and among populations that are homeless or reside in informal settlements.

Among a largely illiterate study population, continued participation relied crucially on oral recruitment instructions. Most participants remembered the oral instructions well and were able to share them with their recruits, though recruits came by the office occasionally to ask questions and inquire about the validity of their coupons. One data collector served as the first point of contact for participants and potential participants. She managed all inquiries, along with the informed consent process and linguistic matching of participants with data collectors. During opening hours, this process ran quite smoothly. However, there was one instance in which two young girls misunderstood the recruitment instructions from their recruiter. These participants appeared at the office to participate in the early morning and reportedly waited around for two hours until the office opened. In the last days of data collection, data collectors

informed participants that the study was concluding at the end of the week. The office remained open for one week after the study ended in order to pay out recruitment bonuses and answer any questions about the study, but only one participant returned to collect her recruiting bonus. No potential participants attempted to redeem expired coupons once the study ended, suggesting that the oral participation instructions were well received and understood.

> **Lesson learned:** When conducting RDS among a largely illiterate population, keep recruiting instructions concise. Encourage recruiters and recruits to visit the study site if they forget oral participation details. Establish a member of the research team at the place of interview as a visible first point of contact for participants and potential participants seeking additional information.

3.3.2 Kayayei data collectors

Hiring and training kayayei with a high school education to work as data collectors was a valuable recommendation from the formative assessment period. The kayayei data collectors were integral to the success of this research. They minimized participants' fears of participating in research and put participants at ease. Data collectors repeatedly mentioned that kayayei said they participated because data collectors could speak their languages and identify with the group. Data collectors also reported multiple participants saying that they only spoke truthfully since the data collector spoke the same language. In addition, kayayei data collectors were well positioned to identify ineligible participants (e.g., beggars) who might have attempted to participate using fabricated responses to obtain the incentive.

> **Lesson learned:** Identify appropriate data collectors with input from the study's target population, especially when populations have reported negative prior experiences with researchers, outsiders, and census enumerators. Hiring data collectors who identify as insiders can help the study population feel more at ease and minimize social distance. For surveys of a sensitive nature, gender matching may further encourage rapport and free discussion between participants and data collectors (Oakley 1981).

3.3.3 Incentives

Incentives are increasingly used in research to motivate and recruit participants. The potential of incentives to coerce or to exert undue influence on participation is understandably controversial, particularly in studies where the research is risky and degrading (Grant and Sugarman 2004; Singer and Couper 2008). For minimal-risk, nondegrading human-subjects research that fulfils the usual ethical criteria, like this study, incentives often pose no ethical problems (Grant and Sugarman 2004; Singer and Couper 2008).

In practice, the initial five-cedi participation incentive operated as compensation for the actual and potential missed work that participants could be expected to forego in order to participate in this study. To make participation worthwhile, the participation incentive had to compensate participants for their time spent participating in the study (2 cedis) as well as for the time involved in traveling to the office and any potential transportation expenses (3 cedis). This incentive did not appear to be excessive. Only five ineligible people attempted to participate, and the data collection team heard no reports of people selling or bartering coupons. Data collectors did not identify any repeat participants. When asked to evaluate the quality of information that participants shared, data collectors assessed 97.0% of participants as providing high quality information, 2.1% as providing fair quality information, and 1.0% as providing low quality information. Data collectors reported that participants assessed as providing fair or low quality information tended to be shy, did not know their exact age, or needed to have questions repeated. No data collectors reported that they suspected the incentive affected participants' responses, leading to lower quality data.

> **Lesson learned:** Compensating participants for their time completing the survey may insufficiently motivate working individuals to participate. The participation incentive may need to include compensation for time traveling to the office and transportation expenses. The formative research phase should address these factors.

To ensure continued recruitment, the maximum recruitment bonus remained at 3 cedis throughout the study. Participants receiving three coupons would receive 1 cedi per recruit, those with two recruiting coupons would be paid 1.5 cedis per recruit, and those with one recruiting coupon would be paid 3 cedis for successful recruitment. Of the 575 participants who received recruiting coupons, 68.4% returned to the office to collect payment for successfully recruiting additional participants. It is possible that those who did not return for their recruiting bonuses migrated elsewhere, deemed the bonus too low to justify time away from work to collect it, or were waiting for confirmation from their recruits that their recruits had completed the study.

Participants also identified nonfinancial reasons for participation. They reported participating because the research team treated them with respect that they seldom received in Accra:

> You see us to be human beings, like you. […] How they [in Accra] open their mouths to insult us, they don't see us to be human beings like the way you see us to be human beings. (Participant, from the Upper East Region, 20s)

Others valued the opportunity to have their voices heard and to advocate for increased development and economic opportunities in northern Ghana. Whilst the indepth interviews focused on participants' migration histories and their use of health services for recent illness/injury, interview respondents frequently discussed their motivations for participating in the survey research. Many participants were hopeful that the findings from this research would translate into help:

> God give you people more knowledge so that you people will help us at the north. (Participant and RDS seed, from the Northern Region, 20s)

> Now when they are telling us what is the problem, they know that it will go outside to USA or London so that they [outsiders] will know their [kayayei's] problems. (Translator, from the Northern Region, 30s)

The theme of beneficence emerged among respondents who hoped that the study would benefit their families and their communities. Some participants asked for assistance in paying the school fees of their children or siblings, for example. Whilst financial assistance was not provided by this research project, conducting the study out of the KYA office enabled the research team to raise awareness of local initiatives by nongovernmental organizations and to link participants with appropriate, albeit limited, services.

> **Lesson learned:** During the formative research phase, gather information about local initiatives and services that may be relevant to the study population. Have details of these services available during the study to assist participants in connecting with necessary services.

4. Discussion

Sampling kayayei migrants using RDS required a nuanced understanding of both the participants' experiences and the local community to ensure successful implementation and sufficient recruitment. Compared to other studies collecting data on kayayei, this study generated the most comprehensive dataset to date. It overcame many challenges and limitations reported in previous studies (Table 5) and permitted the collection of more data in a shorter timeframe.

Prior studies on kayayei report concerns with the representativeness of their samples (Opare 2003; Yeboah 2010; Addai 2011; Oberhauser and Yeboah 2011; Agyei, Kumi, and Yeboah 2016). To improve representativeness in this study and minimize known sources of nonparticipation bias, the formative research phase discussed past studies' sampling challenges (Table 5) and potential solutions. By using RDS rather than a sampling frame generated from phone numbers, for example, this study included 64.8% of respondents who would have been excluded from prior research (Yiran, Teye, and Yiran 2015). Instead of requiring that participants speak Twi (Addai 2011), kayayei data collectors accommodated participants in multiple dialects. The kayayei data collectors also facilitated greater participation among individuals who were suspicious of outsiders, reducing that barrier to participation (Baah-Ennumh, Amponsah, and Adoma 2012).

By design, RDS minimized prior challenges regarding participant availability and willingness to take time away from work to engage in extended conversations (Opare 2003; Awumbila and Ardayfio-Schandorf 2008; Yeboah 2010). This study remained unaffected by the high degree of mobility that challenged others' efforts to track participants and to collect data (Opare 2003; Yeboah and Appiah-Yeboah 2009; Yeboah 2010; Addai 2011), since participation required visiting a central office. Basing the research in a central office will have restricted potential recruits working and living outside the central business district; however, the surveys captured participants living in 29 communities and working in 44 places within Accra.

> **Lesson learned:** For widely dispersed and highly mobile populations, multiple field sites will minimize participants' need for transportation and will better capture their geographical distribution. If financial and personnel resource constraints make such an endeavor unfeasible, select the most appropriate site in close collaboration with stakeholders, including members of the target population.

Whilst the data from this study permits stronger generalizations and population-level inferences than those from prior studies employing nonrandom sampling methods, not all results are generalizable. Despite discussing languages and cross-group

recruitment during the formative research phase, this study did not obtain equilibrium for ethnic group or the closely related variables religion, region of birth, and place of residence in Accra. Such data cannot be generalised to the larger kayayei population. For the majority of kayayei migrants, linguistic differences between ethnic groups likely restricted cross-group recruitment and equilibrium. Though the survey did not capture all languages a participant spoke, the survey recorded the language of the interview. Ultimately, a minimum of one in seven participants spoke at least two languages. English- and Twi-speaking participants would have been well positioned to recruit across ethnic group, as would other multilingual participants who the survey questionnaire could not identify. It is possible that the sample's bias towards well-connected respondents also restricted cross-group recruitment and equilibrium. Generalizations are stronger for variables pertaining to health (e.g., ever given birth, ever had a health problem caused by kayayoo work) and migration behaviours (e.g., send remittance(s), recommend kayayoo work to friend in the north) that obtained equilibrium.

> **Lesson learned:** It is unclear whether participants' preference to recruit within ethnic group resulted from linguistic barriers, from a preference to share resources (i.e., the RDS incentives) internally, and/or from other factors. If this preference results from linguistic barriers, more than one in seven members of the target population would likely need to identify as bilingual or multilingual to achieve equilibrium for the variable ethnic group. Future studies involving migrant populations comprised of multiple ethnolinguistic groups should consider capturing participants' languages in order to better assess the impact of linguistic differences on delaying or preventing achievement of equilibrium. Whilst RDS routinely collects nonresponse data from participants returning for their recruiting bonuses, researchers could also ask returning participants why they selected the recruit(s) to whom they gave their coupons, to uncover recruiting barriers and preferences.

Table 5:   **Sampling challenges/limitations among studies of kayayei in Ghana and how RDS performed against these approaches**

Author (year)	Study date(s)	Kayayei population	Data collection tools	Sampling method	Reported sampling challenges/limitations	How RDS performed against this approach
(Addai 2011)	July to Sept. 2009	209 *kayayei* migrants in Kumasi	Survey	Key informant sampling using *kayayei* fluent in Twi to survey respondents	Concerns with sample representativeness; unreliable/ unknown population size preventing mathematical determination of sample size; no residential addresses from which to sample	Larger sample collected in a shorter time frame; sampling frame unnecessary; multilingual data collectors surveyed *kayayei* in northern dialects; sampling estimates may be generated in advance to assist in study planning and implementation; RDS produces a weighted sample proven to be unbiased for samples of meaningful size; nonresponse data available
(Agyei, Kumi et al. 2016)	July and Aug. 2014	45 *kayayei* in Makola and Agbogbloshie Markets, Accra	Semi-structured questionnaires, interviews	Mix of snowball sampling and convenience sampling relying upon assistance of local leaders	Concerns with sample representativeness; unreliable/unknown population size preventing mathematical determination of sample size; overrepresentation of one ethnic or linguistic group	Larger sample collected in a similar time frame; multilingual data collectors accommodated participants in seven different languages; sampling estimates may be generated in advance to assist in study planning and implementation; RDS produces a weighted sample proven to be unbiased for samples of meaningful size; nonresponse data available
(Awumbila and Ardayfio-Schandorf 2008)	2004 to 2005 (survey data)	100 young female porters or *kayayei* in the central business district of Accra	Questionnaires, key informant interviews (KIIs), case studies, life histories, focus group discussions (FGDs), participant observation	Purposive sampling of working participants	Participant availability and willingness to take time away from work	Participants willing to take time away from work when given control over the timing of their participation and an incentive recognizing missed work; larger sample collected in a shorter time frame; nonresponse data available

Table 5: **(Continued)**

Author (year)	Study date(s)	Kayayei population	Data collection tools	Sampling method	Reported sampling challenges/limitations	How RDS performed against this approach
(Baah-Ennumh, Amponsah, and Adoma 2012)	–	100 female head porters in Kumasi	Semi-structured questionnaires and interview guides	'Intuitive' selection of participants using snowball sampling	Lack of a sampling frame; unreliable/unknown population size preventing mathematical determination of sample size; population's unwillingness to communicate with strangers out of suspicion	Sampling frame unnecessary; peer-recruitment enabled recruits to discuss study with known recruiter before participating, minimizing potential fears of research; sampling estimates may be generated in advance to assist in study planning and implementation; nonresponse data available
(Oberhauser and Yeboah 2011)*	Six months in 2005	Survey of 80 female and 40 male porters in Accra; further IDIs with 15 female and 10 male porters	Survey containing semi-structured and structured questions, IDIs, FGDs	Snowball sampling	Concerns with sample representativeness	Larger sample collected in a shorter time frame; RDS produces a weighted sample proven to be unbiased for samples of meaningful size; nonresponse data available
(Opare 2003)	–	700 young women and teenage girls	Unstructured interviews (in groups and exclusive), personal observation	Nonrandom convenience sample at markets and places where kayayei take breaks	Participant availability and willingness to take time away from work; lack of a sampling frame due to highly mobile population; concerns with sample representativeness	Participants willing to take time away from work when given control over the timing of their participation and an incentive recognizing missed work; sampling frame unnecessary; RDS produces a weighted sample proven to be unbiased for samples of meaningful size; nonresponse data available
(Shamsu-Deen 2013)	June 2011 to Jan. 2012	400 female porters in Accra	Structured questionnaire, FGDs, IDIs, life histories, case studies, KIIs	Purposive sampling in selected markets	Unreported	Larger sample collected in a shorter time frame; nonresponse data available
(Tufeiru 2014)	–	15 migrant women porters in Accra	Semi-structured interviews	Purposeful sampling based on registered members of the Kayayo Youth Association	Unreported	Nonresponse data available

Table 5: **(Continued)**

Author (year)	Study date(s)	*Kayayei* population	Data collection tools	Sampling method	Reported sampling challenges/limitations	How RDS performed against this approach
(Wilson and Mittelmark 2013)	–	8 female porters aged 20-25 years from the north with ≥ 6 months of porter experience	Interviews and observations	A purposive sampling technique combined with snowball sampling	Sampling method did not produce intended sample of well-adjusted participants	Nonresponse data available
(Yeboah 2010)*	Mid-June to mid-Dec. 2005	Survey of 80 female and 40 male porters in Accra; further IDIs with 15 female and 10 male porters	Survey, IDIs, observations, field notes, FGDs	Convenience sampling at four selected sites in Accra	Lack of a sampling frame due to highly mobile population; participant availability and willingness to take time away from work; potential bias from convenience sampling; nongeneralizable data; concerns with sample representativeness	Participants willing to take time away from work when given control over the timing of their participation and an incentive recognizing missed work; larger sample collected in a shorter time frame; RDS produces a weighted sample proven to be unbiased for samples of meaningful size; nonresponse data available
(Yeboah and Appiah-Yeboah 2009)*	Mid-June to mid-Dec. 2005	Survey of 80 female and 40 male porters in Accra; further IDIs with 15 female and 10 male porters	Survey questionnaires, IDIs, observations, FGDs	Primarily snowball sampling	Lack of a sampling frame due to highly mobile population; nongeneralizable data	Sampling frame unnecessary; larger sample collected in a shorter time frame; nonresponse data available
(Yiran, Teye, and Yiran 2015)	–	70 female head porters at two markets in Accra who were pregnant or nursing	Survey questionnaire, IDIs	Ad hoc listing exercise collected *kayayei's* phone numbers; researchers selected participants from this sampling frame using simple random sampling; administered questionnaires in northern languages	Unreliable/unknown population size preventing mathematical determination of sample size	Mobile phone ownership was not a prerequisite for participating in this study; sampling estimates may be generated in advance to assist in study planning and implementation; nonresponse data available

* Analyses using this sample are reported in the following studies: Oberhauser and Yeboah 2011, Yeboah 2010, and Yeboah and Appiah-Yeboah 2009.

Several sources of potential bias or imprecision affect RDS and this data. In addition to the previously mentioned methodological limitations and assumptions, RDS estimates are particularly sensitive to errors in reported PNS (Mills et al. 2014). This study refined the PNS question in an attempt to minimize such errors, and data collectors encouraged participants to spend time counting specific individuals in their networks before answering PNS questions. Participants are supposed to recruit both strong and weak ties, but the majority of participants (96.3%) reported knowing their recruiter 'very well.' The final sample appears biased towards well-connected respondents, which may lead to biased population proportion estimates. Yet among participants who knew their recruiter for three weeks or less ($n = 35$), 57.1% of participants reported knowing these new contacts 'very well.' Participants may have inflated the strength of their recruiter-recruit relationships, leading to an overestimation of strong ties.

Nonresponse bias and measurement errors also affect this survey data. Whilst participants returning to collect their recruitment payments provided nonresponse data for 16 potential recruits who declined coupons, 302 distributed recruitment coupons (33.0%) were unredeemed. Having no time to participate, fearing research, and lacking interest in the study contributed in part to nonresponses; however, the actual number of nonresponses is likely much lower than the number of unredeemed coupons. Informal discussions with participants revealed that coupon theft, coupon deterioration, and return migrations contributed to nonresponse. Quality checks built into the survey and participant flow process uncovered measurement errors, specifically nonsampling errors, that occurred on several known occasions when data collectors and respondents misunderstood survey questions. These errors were immediately discussed with the research team and corrected before the participant departed.

5. Conclusion

This study expands the evidence base for the use of RDS among migrant populations in LMIC and demonstrates that it is feasible to quickly and efficiently collect extensive data from a large sample of migrants. RDS will not be appropriate for all migrant populations. The migrant kayayei are a clearly defined social group, who recognize one another as part of the group; however, kayayei come from multiple northern tribes. Ethnic ties remained strong for most participants. In countries with numerous ethnic groups and dialects, researchers may struggle to recruit a population of internal migrants defined by factors such as occupation or health condition. Ethnolinguistic differences between internal migrants may result in recruiting preferences or barriers that restrict some respondents from participating, thus delaying or preventing

achievement of equilibrium. In such cases, researchers should consider alternative sampling methods or should consider narrowing their population to a group with ethnolinguistic overlap. Implementing RDS among a specific ethnolinguistic migrant group would likely facilitate achieving equilibrium with fewer waves. It is possible that these barriers may be weaker when implementing RDS among international migrants who originate from the same country if national identity supersedes intra-national identity, helping facilitate recruitment across ethnic groups.

Researchers studying migrants in LMIC should consider RDS if formative research indicates that the method is likely to be appropriate and well received by the study population. Formative research with members of the target population, community leaders, and key stakeholders was integral to the successful implementation of RDS in this study. This research included assessing logistical factors preimplementation (e.g., appropriate incentives, coupon design, whether personal networks are sufficient); however, the formative research also expanded the logistics of implementing RDS in order to identify specific factors regarding both the population and the place that could potentially affect implementation (e.g., kayayei are resistant to speak with outsiders given past experiences with census enumerators and talk of forcibly sending kayayei back north; participants may need to bring children with them in order to participate).

RDS may allow researchers in LMIC to stretch limited resources, as it has been found to be faster, less expensive, and easier to implement than other methods for examining hard-to-reach populations (Semaan, Lauby, and Liebman 2002; Salganik and Heckathorn 2004). Future studies implementing RDS among migrant populations in such settings should include members of the study population as data collectors and as advisors throughout the study to advise on the appropriateness and feasibility of implementing RDS. Although nonfinancial incentives may motivate potential recruits to participate in the study, incentives must recognize participants' time, including participation as well as travel to and from the study site. In settings where a high rate of coupon deterioration may occur, coupon design must consider this factor.

Employing members of the target population to collect data was critical in building trust in the research. It ultimately resulted in high-quality data that would likely have been more difficult to collect with data collectors who identified as outsiders. By incorporating these accommodations into the RDS implementation process there was no need to significantly adapt RDS to maximize the sample, as others have done (Platt, Luthra, and Frere-Smith 2015). This research achieved a large sample size (n = 625) with a relatively small team (n = 8) and budget. RDS quickly and efficiently produced the most comprehensive dataset on a migrant population, which conventional survey and census efforts have struggled to capture. The lessons learned from this case study illustrate that whilst RDS it is not a one-size-fits-all solution for sampling hard-to-reach

migrants in LMIC, it can be a powerful tool to uncover and recruit hard-to-reach migrant populations.

6. Acknowledgements

References

Addai, I. (2011). Estimating remittances in the informal sector labour market in a developing economy: A micro-level evidence on Kayayoo migrants in Kumasi, Ghana. *The Social Sciences* 6(4): 313–317. doi:10.3923/sscience. 2011.313.317.

Agadjanian, V. and Zotova, N. (2012). Sampling and surveying hard-to-reach populations for demographic research: A study of female labor migrants in Moscow, Russia. *Demographic Research* 26(5): 131–150. doi:10.4054/Dem Res.2012.26.5.

Agyei, Y.A., Kumi, E., and Yeboah, T. (2016). Is better to be a kayayei than to be unemployed: Reflecting on the role of head portering in Ghana's informal economy. *GeoJournal* 81(2): 293–318. doi:10.1007/s10708-015-9620-z.

Anarfi, J.K. and Appiah, M. (2009). The phenomenon of independent child migration in Ghana in the context of a globalised world. In: Anarfi, J.K. and Kwankye, S.O. (eds.). *Independent migration of children in Ghana*. Legon: Sundel Services: 45–70.

Awumbila, M. and Ardayfio-Schandorf, E. (2008). Gendered poverty, migration and livelihood strategies of female porters in Accra, Ghana. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography* 62(3): 171–179. doi:10.1080/00291950802335772.

Baah-Ennumh, T.Y., Amponsah, O., and Adoma, M.O. (2012). The living conditions of female head porters in the Kumasi metropolis, Ghana. *Journal of Social and Development Sciences* 3(7): 229–244.

Bryant, J. (2014). Using respondent-driven sampling with 'hard to reach' marginalised young people: Problems with slow recruitment and small network size. *International Journal of Social Research Methodology* 17(6): 599–611. doi:10.1080/13645579.2013.811921.

Carballo-Diéguez, A., Balan, I., Marone, R., Pando, M.A., Dolezal, C., Barreda, V., Leu, C.-S., and Ávila, M.M. (2011). Use of respondent driven sampling (RDS) generates a very diverse sample of men who have sex with men (MSM) in Buenos Aires, Argentina. *PLoS One* 6(11): e27447. doi:10.1371/ journal.pone.0027447.

Crush, J. and Tawodzera, G. (2014). Medical xenophobia and Zimbabwean migrant access to public health services in South Africa. *Journal of Ethnic and Migration Studies* 40(4): 655–670. doi:10.1080/1369183X.2013.830504.

Daily Express (2010, September 28). "Kayayei" numbers higher than anticipated. *Daily Express*. http://statsghana.gov.gh/docfiles/news/Kayayei.pdf.

Daily Graphic (2010, September 28). Hectic time for census teams. *Daily Graphic*. https://www.modernghana.com/news/298213/hectic-time-for-census-teams.html.

Day, C. (2014). Giving the vulnerable a voice: Ethical considerations when conducting research with children and young people. In: Lunn, J. (ed.). *Fieldwork in the global South: Ethical challenges and dilemmas*. London: Routledge: 192–205.

Decker, M.R., Marshall, B.D., Emerson, M., Kalamar, A., Covarrubias, L., Astone, N., Wang, Z., Gao, E., Mashimbye, L., Delany-Moretlwe, S., Acharya, R., Olumide, A., Ojengbede, O., Blum, R.W., and Sonenstein, F.L. (2014). Respondent-driven sampling for an adolescent health study in vulnerable urban settings: A multi-country study. *Journal of Adolescent Health* 55(6S): S6–S12. doi:10.1016/j.jadohealth.2014.07.021.

Driedger, L. and Peters, J. (1977). Identity and social distance: Towards understanding Simmel's 'The Stranger.' *Canadian Review of Sociology* 14(2): 158–173. doi:10.1111/j.1755-618X.1977.tb00339.x.

Evans, R. and Becker, S. (2009). *Children caring for parents with HIV and AIDS: Global issues and policy responses*. Bristol: Policy Press.

Friberg, J.H. and Horst, C. (2014). RDS and the structure of migrant populations. In: Tyldum, G. and Johnston, L. (eds.). *Applying respondent driven sampling to migrant populations: Lessons from the field*. London: Palgrave Pivot: 17–26.

Gile, K.J. and Handcock, M.S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology* 40(1): 285–327. doi:10.1111/j.1467-9531.2010.01223.x.

Goel, S. and Salganik, M.J. (2009). Respondent-driven sampling as Markov chain Monte Carlo. *Statistics in Medicine* 28(17): 2202–2229. doi:10.1002/sim.3613.

Goel, S. and Salganik, M.J. (2010). Assessing respondent-driven sampling. *Proceedings of the National Academy of Sciences* 107(15): 6743–6747. doi:10.1073/pnas.1000261107.

Grant, R.W. and Sugarman, J. (2004). Ethics in human subjects research: Do incentives matter? *Journal of Medicine and Philosophy* 29(6): 717–738. doi:10.1080/03605310490883046.

GSS, GHS, and ICF Macro (2009). Ghana Demographic and Health Survey 2008. Calverton and Accra: Ghana Statistical Service, Ghana Health Service, and ICF Macro. https://dhsprogram.com/publications/publication-fr221-dhs-final-reports.cfm.

Heckathorn, D. (2007). Extensions of respondent driven sampling: Analyzing continuous variables and controlling for differential recruitment. *Sociological Methodology* 37(1): 151–207. doi:10.1111/j.1467-9531.2007.00188.x.

Heckathorn, D.D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems* 44(2): 174–199. doi:10.2307/3096941.

Heckathorn, D.D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems* 49(1): 11–34. doi:10.1525/sp.2002.49.1.11.

International Organization for Migration (2010). *Integrated Biological and Behavioural Surveillance Survey among migrant female sex workers in Nairobi, Kenya*. Nairobi: International Organization for Migration.

Johnston, L., Rodriguez, L., and Napierala, J. (2014). Measuring personal network size in RDS. In: Tyldum, G. and Johnston, L. *Applying respondent driven sampling to migrant populations: Lessons from the field*. London: Palgrave Pivot: 27–36.

Johnston, L., Thurman, T., Mock, N., Nano, L., and Carcani, V. (2010). Respondent-driven sampling: A new method for studying street children with findings from Albania. *Vulnerable Children and Youth Studies* 5(1): 1–11. doi:10.1080/17450120903193923.

Kearney, H. (2013, November 22). Ghana's female porters get family planning advice. *Guardian*. https://www.theguardian.com/global-development-professionals-network/2013/nov/22/ghanas-female-porters-get-family-planning-advice.

Keygnaert, I., Dialmy, A., Manco, A., Keygnaert, J., Vettenburg, N., Roelens, K., and Temmerman, M. (2014). Sexual violence and sub-Saharan migrants in Morocco: A community-based participatory assessment using respondent driven sampling. *Globalization and Health* 10(32). doi:10.1186/1744-8603-10-32.

Khamsiriwatchara, A., Wangroongsarb, P., Thwing, J., Eliades, J., Satimai, W., Delacollette, C., and Kaewkungwal, J. (2011). Respondent-driven sampling on the Thailand–Cambodia border: I. Can malaria cases be contained in mobile migrant workers? *Malaria Journal* 10(120): 1–11. doi:10.1186/1475-2875-10-120.

Kubal, A., Shvab, I., and Wojtynska, A. (2014). Initiation of the RDS recruitment process: Seed selection and role. In: Tyldum, G. and Johnston, L. *Applying respondent driven sampling to migrant populations: Lessons from the field*. London: Palgrave Pivot: 37–48. doi:10.1057/9781137363619_5.

Kwankye, S.O. and Addoquaye Tagoe, C. (2009). City life outside the home: The experiences of independent child migrants in Ghana. In: Anarfi, J.K. and Kwankye, S.O. (eds.). *Independent migration of children in Ghana*. Legon: Sundel Services.

Lattof, S.R. (2018). Health insurance and care-seeking behaviours of female migrants in Accra, Ghana. *Health Policy and Planning*. doi:10.1093/heapol/czy012.

Liu, H., Li, J., Ha, T., and Li, J. (2012). Assessment of random recruitment assumption in respondent-driven sampling in egocentric network data. *Social Networking* 1(2): 13–21. doi:10.4236/sn.2012.12002.

Mills, H.L., Johnson, S., Hickman, M., Jones, N.S., and Colijn, C. (2014). Errors in reported degrees and respondent driven sampling: Implications for bias. *Drug and Alcohol Dependence* 142: 120–126. doi:10.1016/j.drugalcdep.2014.06.015.

Montealegre, J., Risser, J., Selwyn, B., McCurdy, S., and Sabin, K. (2013). Effectiveness of respondent driven sampling to recruit undocumented Central American immigrant women in Houston, Texas for an HIV behavioral survey. *AIDS and Behavior* 17(2): 719–727. doi:10.1007/s10461-012-0306-y.

Oakley, A. (1981). Interviewing women: A contradiction in terms. In: Roberts, H. (ed.). *Doing feminist research*. London: Routledge: 31–61.

Oberhauser, A.M. and Yeboah, M.A. (2011). Heavy burdens: Gendered livelihood strategies of porters in Accra, Ghana. *Singapore Journal of Tropical Geography* 32(1): 22–37. doi:10.1111/j.1467-9493.2011.00417.x.

Okal, J., Raymond, H.F., Tun, W., Musyoki, H., Dadabhai, S., Broz, D., Nyamu, J., Kuria, D., Muraguri, N., and Geibel, S. (2016). Lessons learned from respondent-driven sampling recruitment in Nairobi: Experiences from the field. *BMC Research Notes* 9(1): 1–13. doi:10.1186/s13104-016-1965-y.

Opare, J.A. (2003). Kayayei: The women head porters of southern Ghana. *Journal of Social Development in Africa* 18(2): 33–48.

Parliament of the Republic of Ghana (2016). Parliamentary debates: Official report, Friday, 4th March, 2016. *Parliamentary Debates* 83(24): 1889–2072.

Platt, L., Luthra, R., and Frere-Smith, T. (2015). Adapting chain referral methods to sample new migrants: Possibilities and limitations. *Demographic Research* 33(24): 665–700. doi:10.4054/DemRes.2015.33.24.

Rudolph, A.E., Fuller, C.M., and Latkin, C. (2013). The importance of measuring and accounting for potential biases in respondent-driven samples. *AIDS and Behavior* 17(6): 2244–2252. doi:10.1007/s10461-013-0451-y.

Salganik, M. (2006). Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *Journal of Urban Health* 83(1): 98–112. doi:10.1007/s11524-006-9106-x.

Salganik, M.J. and Heckathorn, D.D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology* 34(1): 193–239. doi:10.1111/j.0081-1750.2004.00152.x.

Salifu, M.I. (2015). Unpublished oral interview with Mohammed Salifu, Director and Founder of Kayayei Youth Association of Ghana. Conducted by S.R. Lattof.

Semaan, S., Lauby, J., and Liebman, J. (2002). Street and network sampling in evaluation studies of hiv risk-reduction interventions. *AIDS Review* 4: 213–223.

Shamsu-Deen, Z. (2013). Migrattion [sic] and health nexus: A case of female porters (Kayayei) in Accra, Ghana. *Research on Humanities and Social Sciences* 3(3): 103–109.

Singer, E. and Couper, M.P. (2008). Do incentives exert undue influence on survey participation? Experimental evidence. *Journal of Empirical Research on Human Research Ethics* 3(3): 49–56. doi:10.1525/jer.2008.3.3.49.

Spiller, M.W., Cameron, C., and Heckathorn, D.D. (2012). *RDS Analysis Tool 7.1: User manual*. Ithaca: Cornell University.

Tufeiru, A. (2014). The nexus of female capital and north–south labor migration in Ghana: A potential remedy from microfinance. *Journal of Developing Societies* 30(1): 91–114. doi:10.1177/0169796X13504680.

Tyldum, G. and Johnston, L. (eds.) (2014). *Applying respondent driven sampling to migrant populations: Lessons from the field*. London: Palgrave Pivot. doi:10.1057/9781137363619.

Tyldum, G., Rodriguez, L., Bjørkhaug, I., and Wojtynska, A. (2014). Deciding on and distributing incentives in RDS. In: Tyldum, G. and Johnston, L. (eds.). *Applying respondent driven sampling to migrant populations: Lessons from the field*. London: Palgrave Pivot.

UNDESA (2013). Cross-national comparisons of internal migration: An update on global patterns and trends. New York: United Nations Department of Economic and Social Affairs.

United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. Geneva: United Nations.

Volz, E. and Heckathorn, D.D. (2008). Probability based estimation theory for respondent driven sampling. *Journal of Official Statistics* 24(1): 79–97.

Volz, E., Wejnert, C., Cameron, C., Spiller, M., Barash, V., Degani, I., and Heckathorn, D.D. (2012). Respondent-driven sampling analysis tool (RDSAT). Ithaca: Cornell University. Version 7.1.

Wejnert, C., Pham, H., Krishna, N., Le, B., and DiNenno, E. (2012). Estimating design effect and calculating sample size for respondent-driven sampling studies of injection drug users in the United States. *AIDS and Behavior* 16(4): 797–806. doi:10.1007/s10461-012-0147-8.

White, R.G., Hakim, A.J., Salganik, M.J., Spiller, M.W., Johnston, L.G., Kerr, L., Kendall, C., Drake, A., Wilson, D., Orroth, K., Egger, M., and Hladik, W. (2015). Strengthening the reporting of observational studies in epidemiology for respondent-driven sampling studies: 'STROBE-RDS' statement. *Journal of Clinical Epidemiology* 68(12): 1463–1471. doi:10.1016/j.jclinepi.2015.04.002.

Wilson, A. and Mittelmark, M.B. (2013). Resources for adjusting well to work migration: Women from northern Ghana working in head porterage in Greater Accra. *Africa Today* 59(4): 24–38. doi:10.2979/africatoday.59.4.25.

Ye, Y., Wamukoya, M., Ezeh, A., Emina, J.B.O., and Sankoh, O. (2012). Health and demographic surveillance systems: A step towards full civil registration and vital statistics system in sub-Sahara Africa? *BMC Public Health* 12(1): 741. doi:10.1186/1471-2458-12-741.

Yeboah, M.A. (2010). Urban poverty, livelihood, and gender: Perceptions and experiences of porters in Accra, Ghana. *Africa Today* 56(3): 42–60. doi:10.2979/aft.2010.56.3.42.

Yeboah, M.A. and Appiah-Yeboah, K. (2009). An examination of the cultural and socio-economic profiles of porters in Accra, Ghana. *Nordic Journal of African Studies* 18(1): 1–21.

Yiran, G.-S., Teye, J., and Yiran, G.B. (2015). Accessibility and utilisation of maternal health services by migrant female head porters in Accra. *Journal of International Migration and Integration* 16(4): 929–945. doi:10.1007/s12134-014-0372-2.