Feasibility of using respondent-driven sampling to recruit participants in superdiverse neighbourhoods for a general health survey

Florence Samkange-Zeeb, Ronja Foraita, Stefan Rach, Tilman Brand

This is the author's final accepted version. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Feasibility of using respondent-driven sampling to recruit participants in superdiverse neighbourhoods for a general health survey

Abstract

Objectives

Respondent driven sampling (RDS), a modified chain-referral system, has been proposed as a strategy for reaching 'hidden' populations. We applied RDS to assess its feasibility to recruit 'hard-to-reach' populations such as migrants and the unemployed in a general health survey and compared it to register-based sampling (RBS).

Methods

RDS was applied parallel to standard population RBS in two superdiverse neighbourhoods in Bremen, Germany. Prevalences of sample characteristics of interest were estimated in RDS Analyst using the Successive Sampling estimator. These were then compared between the samples.

Results

Only 115 persons were recruited via RDS compared to 779 via RBS. The prevalence of (i) migrant background, (ii) unemployment, and (iii) poverty risk was significantly higher in the RDS than in the RBS sample. The respective estimates were (i) 51.6 % vs 32.5% (95% CI_{RDS} 40.4–62.7), (ii) 18.1% vs 7.5 % (95% CI_{RDS} 8.4–27.9) and (iii) 55.0% vs 30.4% (95% CI_{RDS} 41.3–68.7).

Conclusions

Although recruitment was difficult and the number of participants was small, RDS proved to be a feasible method for reaching migrants and other disadvantaged persons in our study.

Introduction

Although migrants make up quite a substantial proportion of the population in many countries (United Nations 2016), they are often underrepresented in health research (Weber et al. 2009; Redwood and Gill 2013; Quay et al. 2017). This not only limits the validity and generalizability of findings, but also negatively impacts on health care provision equity (George, Duran and Norris 2014). Register-based sampling (RBS) is cited as one of the main contributors towards the lower participation of persons with a migrant background, due to difficulties in identifying migrants and language barriers (Reiss et al. 2014). To try and combat this, different methods, including snowball sampling, targeted-sampling as well as Respondent-Driven Sampling (RDS), have been put forward (Shaghaghi, Bhopal and Sheikh 2011; Hu, Wong and Wang 2015). RDS is a modified chain-referral sampling approach which was developed by Heckathorn (1997). Recruitment starts with seeds (initial respondents), well-connected persons in the target population, who are purposefully recruited. Each seed is issued with a small number of coupons which they use to recruit further participants. The coupons serve as invitations to the study and also identify the presenting person as a valid recruit. Recruited persons are also issued with coupons and asked to recruit further persons from their own social networks. The coupons have unique identifying numbers used to link the recruiter with the recruited person, thereby enabling the reconstruction of recruitment chains. This also allows RDS to take clustering and similarities among the respondents as well as post hoc sampling probabilities into account during the analysis, which is a great advantage over other chain-referral methods such as snowballing (Salganik 2006; Johnston 2014). Another important aspect of the method is that recruiters receive incentives for each person they recruit for the study.

RDS has been used in numerous studies focusing on "hidden populations", that is, persons belonging to a group whose members are stigmatized or for whom membership is associated with illegal behaviour. Up to now the method has mostly been used to recruit specific populations, i.e. clearly defined groups with a common characteristic e.g. men who have sex with men, those having or at risk of acquiring HIV or children living on the street (Salganik 2006; Johnston et al. 2010)

Where it has been used to recruit migrant populations, the focus was on specific migrant groups (Frere-Smith, Luthra, and Platt 2014; Górny and Napierała 2016). To our knowledge RDS has yet to be used in an epidemiological survey not focusing on a specific disease or health outcome and conducted among the general population. In our study we used RDS to recruit a broader spectrum of 'hard to reach' individuals within the context of a general health survey whose aim was to investigate how people in superdiverse neighbourhoods address their health problems. Superdiverse neighbourhoods are described as urban areas in which population groups from different countries, of different religion, age, sex, gender and with different lengths of residence and legal status, live alongside the local population (Vertovec 2007).

The RDS approach was implemented parallel to a register-based approach. The main objectives of the study were to assess the feasibility of implementing RDS in a general health survey and whether 'hard-to-reach' persons, such as migrants and the unemployed are better reached via RDS than via RBS. Our hypothesis was that compared to the RBS group, the RDS sample would comprise a higher proportion of persons with a lower socioeconomic status, of migrant background and with lower German language competencies, with lower self-reported health and/or lower health literacy.

Methods

This study was part of the UPWEB project, Understanding the Practice and Developing the Concept of Welfare Bricolage, whose aim was to look at how residents of superdiverse neighbourhoods address their health problems (Phillimore et al. 2015).

In this paper we report on the survey conducted in Bremen, Germany, in the neighbourhoods Neustadt and Groepelingen. Both neighbourhoods are characterized by a high level of diversity in terms of nationalities of the residents, length of residence, legal status, age-structure and languages spoken, but differ regarding their degree of deprivation. While Neustadt is popular with upper and middle-income families and individuals, Groepelingen is one of the most deprived of the city's 23 neighbourhoods (Senatorin für Arbeit 2009).

Both RBS and RDS were applied in parallel in the two neighbourhoods. As our aim was to assess the feasibility of the RDS approach, we did not perform pre-survey sample size calculations.
Recruitment was carried out from February to May 2017 and the interviews were conducted face-to-face in rooms that had been specifically leased for the survey in the neighbourhoods. The interviewer team was multi-ethnic and multilingual. At the onset of the interview, all participants were again informed about the study aims and then requested to sign a consent form. The study was approved by the ethics committee of the University of Bremen.

Register-Based Sampling (RBS)

A randomly drawn address file of 6000 persons aged 18 and above, 3000 for each neighbourhood, was obtained from the Bremen population registry office. The dataset was then used to draw random samples of about 150 persons at a time, who were then contacted by letter and requested to take part in the survey. On the reverse side of the letters, which were in German, was a short text about the study and how to get in touch with the study centre in English, French, Turkish, Arabic and Bulgarian. An incentive of 10 Euro was offered. Those willing to participate were requested to contact the study centre in the respective neighbourhood via a toll-free telephone number to make an appointment for

the interview. All non-responders received two letters of reminder and at least one attempt was made to reach 683 persons for whom telephone numbers were available. Further, one door-to-door visit was conducted for about 1700 persons.

Respondent-Driven Sampling (RDS)

Selection of seeds

Before the onset of the survey, a convenience sample of 16 persons was identified as seeds, seven in Groepelingen and nine in Neustadt. The persons had either been involved in the qualitative phase of the study as community researchers and/or interviewees, or were acquainted with study staff. They were informed about the aims of the study and their role. After they had indicated their willingness to take part in the study, they were required to contact the respective study centre to make an appointment for their interview.

Coupons

As recommended in the literature on RDS Johnston (2014), the seeds were each issued with three coupons after being interviewed, which they were to use to recruit further participants. The information on the coupons summarised the purpose of the study, the characteristics of the persons being looked for (aged 18 and above and resident in the respective neighbourhood), the incentives being offered as well as specific neighbourhood interview venue and contact details, in the six study languages. The coupons had serial numbers on them that were used to link the recruited person to the recruiter.

Incentives

As in RBS, RDS participants received 10 Euro after their interview. In addition, RDS participants received another 10 Euro for each successful recruitment, that is, for each person they had recruited who was interviewed.

Only persons who presented with a valid coupon were interviewed. Enrolment and payment of incentives was managed using a database application specifically created for this purpose in Microsoft Access (Reineke et al. 2018). The serial number on the presented coupons was scanned and linked with the ID of the recruiter.

Each recruited person who presented for an interview was instructed about how to recruit further persons using the coupons. As RDS is based on the social network theory, the RDS recruits were informed that they should distribute the coupons to persons known to them personally. Although no deadline was set for the distribution of the coupons, the respondents were told that that time was of essence as the study period was only four months.

4

As RDS recruitment progressed slowly, with only 4 of the seeds and one recruit having been interviewed two weeks into the survey, the number of seeds was increased from 16 to 24, 13 in Groepelingen and 11 in Neustadt. Three of the additional seeds were originally RBS participants who were recruited during the interviews. They reported having a large social-network in response to the survey question on personal network size (see following section) and were prepared to try and recruit further persons from their social networks. The other five were recruited through the researchers' networks. A further step taken to speed up recruitment was to increase the number of coupons issued from three to six.

Sample characteristics

Apart from age and sex, socio-demographic variables collected included migration-related and socio-economic indicators. Migrant background was defined as being born outside the country or having parents or grandparents who were born abroad. Among those who were born abroad, self-reported German language proficiency was assessed on a five-point scale, from very good to very poor. For the analysis the scale was dichotomized into high (very good and good) and low (medium to very poor) language proficiency. Level of education was assessed according to the International Standard Classification of Education (ISCED) group using three categories (low: ISCED 0-2; medium: ISCED 3-4; high: ISCED 5-6). We calculated net equivalent household income according to OECD standards (OECD 2011). Respondents were classified as being at risk of poverty if they earned less than 60% equivalent household income in Germany in 2017 (< 1,064€ per month). In addition, respondents reported their employment status. This information was used to calculate the unemployment rate in the samples (number of unemployed divided by the number of employed plus unemployed).

For the analysis of RDS data it is necessary to calculate the sampling probabilities for each respondent based on their personal network size (PNS) as those with larger networks are more likely to be recruited than those with smaller ones. During the analysis, the latter are weighted higher than the former to prevent over and under representation of particular sub-groups (Johnston, Rodriguez and Napierala 2014) The PNS was assessed based on a two-part question as recommended by . In the first part respondents were asked how many people eligible for this survey they know ("How many adults who live in your neighbourhood do you know?"). Knowing meant that the participant knows at least the name of the other person and the person knows theirs. In the second part participants were asked how many of these people they had seen during the previous four weeks. The latter was then used for the calculation of the sampling probabilities as it seemed more plausible that participants would be able to reach at least this number of persons during the course of the study.

Other variables of interest that indicate vulnerability were: self-reported health (Ware and Sherbourne 1992), health literacy (Pelikan, Röthlin and Ganahl 2014) and perceived discrimination (European Social Survey 2012).

Analysis

Using RDS –Analyst (Handcock, Fellows and Gile 2014), bias due to recruitment homophily was assessed for sex, age (≤50 and >50), migrant background, employment status and poverty risk. Recruitment homophily is defined as the tendency for individuals to recruit those with similar characteristics to themselves (Tyldum and Johnston 2014). It is calculated as the ratio of the number of recruits sharing the same characteristic of interest as their recruiter, in relation to the number that would be expected under the assumption of independence. A value greater than one indicates a higher level of similarity in characteristics, while a value of one means that similarities, respectively ties, are random (Johnston and Luthra 2014). Similar to homophily, seed dependence describes the extent to which the characteristics of the sample reflect those of the seeds rather than of the population of interest. Long recruitment chains generally serve to reduce this source of bias. We used recruitment trees to visualize the composition of the different waves and branches of our sample.

The successive sampling (SS) estimator (Gile 2011) was used to estimate the prevalence of the variables of interest based on the RDS sample. The estimator, which assumes that network connections are completely random and that sampling occurs successively without replacement, requires the input of the underlying population size. To this end statistical data for the years 2015 and 2016 were used (Statistisches Landesamt Bremen a 2015, 2016). Underlying population data were not available for the variables for the variables poverty risk, educational level, interview language, German language proficiency, self-reported health, health literacy and perceived discrimination. Hence comparisons were done based on the observed RBS point estimates and the RDS-based estimates only. RDS-based estimates of the sample characteristics were then compared to the characteristics of the RBS sample and the distribution in the underlying population. Differences were regarded as statistically significant if the RBS point estimator or the proportion in the underlying population did not lie within 95% confidence interval of the RDS-based estimates.

RDS-Analyst also provides post hoc design effect estimates, which indicate the increase in variation of RDS estimates compared to estimates from simple random samples. For our study, this information can be used to assess the loss of precision of the estimator under RDS. High values in the design effect (>2) mean that a substantial increase in sample size is required to achieve the same precision in estimates as with simple random sampling (Salganik 2006; Johnston et al. 2016).

Results

Of the 5299 persons contacted via RBS, 779 persons, three of whom became seeds for the RDS approach, participated in the survey (15% response proportion).

For RDS a total of 310 coupons were distributed during the course of the survey and 91 persons (non-seeds) were recruited. This translates into a response proportion of 29%. Only 13 of the 24 seeds managed to recruit at least one person. Of the 13, one was male; seven were older than 50 years and 2/3 had a migrant background. More than half of the 91 recruited persons belonged to the first 3 generations (56%) of the sample, and 38% managed to successfully recruit a further person. Recruitment was generally more successful in Neustadt than in Groepelingen, with females comprising the majority of the sample in both neighbourhoods (Figure 1).

Figure 1 here

Assessment of bias

Recruitment homophily was lowest for the variable unemployed (1.07). The highest values were observed for poverty risk (1.40) and age (1.41). The respective values for sex and migrant background were 1.25 and 1.12.

Homophily as well as seed dependence among the respondents could also be visually assessed using the recruitment trees in Figures 1 and 2. As women were overrepresented in the sample, clusters of women were observed in the sample. Men also tended to recruit men and seeds with a migrant background were more likely to recruit persons who also had a migrant background. The latter resulted in 5 distinct groups in Groepelingen and one in Neustadt (Figure 2). Further, in both neighbourhoods, participants generally did not recruit across poverty risk sub-groups.

Figure 2 here

Comparison of characteristics of RDS and RBS participants

The proportion of persons with a migrant background, of unemployed persons, and of those at poverty risk was significantly higher in RDS- than in RBS (Table 1). No significant differences between the two groups were observed for sex, age, self-reported health, health literacy, personal network size and educational level of participants.

Comparison of the RBS and RDS samples to the underlying population in the neighbourhoods could only be done for the variables sex, age, migrant background and unemployed. In both RBS and RDS, females were oversampled compared to the actual neighbourhood figures, while no differences were observed for age. The proportion of unemployed persons in the RDS sample was similar to that in the underlying population, while that of persons with a migrant background was significantly higher (Table 1). In the RBS sample, the proportion of the unemployed and those with a migrant background was lower than that in the underlying population. The estimated design effect of RDS values for most of the variables ranged between 1.17 and 2.89, with a few outliers beyond this range (Table 1).

Discussion

To our knowledge this is the first epidemiological survey to use RDS in a non-condition related health survey and to compare the approach to the standard RBS. In contrast to other studies, we did not focus on 'hidden' populations or only on specific migrant groups (Strathdee et al. 2008; Montealegre et al. 2013; Frere-Smith, Luthra and Platt 2014; Górny and Napierała 2016), but aimed to reach 'hard-to-reach' groups in the general population.

Our results show that compared to RBS, RDS can be more effective for recruiting 'hard to reach' populations that are not necessarily 'hidden', for example documented persons with a migrant background, the unemployed, or those at poverty risk.

Although the recruitment success proportion for RDS almost doubled the response proportion for RBS (29% vs 15%), the small RDS sample size is a limiting factor for the study. Even though our main focus was on assessing the feasibility of the RDS approach and not on the sample size, we had hoped to reach a sample size comparable to the RBS sample. As this was a feasibility study, we however did not aim to fulfil the recommendations generally made for RDS studies, that the sample size be twice as large as required under simple random sampling (Salganik 2006). Our RDS sample size was probably influenced by several factors that have been discussed in the literature, such as lack of networks and recency of arrival (Frere-Smith, Luthra and Platt 2014). To try and address these we used persons who had helped us identify interview partners during the qualitative phase of the study as seeds. As the persons were able to identify interview partners for us within a very short space of time, we assumed that they were well-connected in their respective communities. Further, as recommended in the literature (Kubal, Shvab and Wojtynska 2014), we selected a diverse group, including men and women, the young and old as well as migrants and non-migrants. However, as experienced in the study by Frere-Smith, Luthra and Platt (2014), in which, according to the authors, six months into the study only five per cent of the 164 participants were referrals and not seeds, recruitment in our study took off very slowly. At study end almost half of our seeds had not managed to recruit another person, raising questions about their suitability. The qualitative interviews on the other hand were conducted at the

interviewee's home or a place of their choice. Although the study centres were easy to reach by public transport, transport costs and time aspects might have reduced willingness to participate. One possible explanation could be that some had difficulties to convince their contacts to participate in the survey as this required the respondents to go to the study centres for the interviews. The fact that the survey did not focus on a specific health outcome might also have led to reluctance to participate. This issue also arose during the qualitative phase of the project, where some people were initially reluctant to participate, stating that they did not have any health concern. They eventually agreed to participate after it had been explained to them that the project aimed to investigate access to healthcare and health-seeking behaviour in general. Having a specific outcome or focusing on a particular population group might have helped raise interest in the survey.

Compared to similar studies on 'hard-to-reach' populations, in which recruitment lasted 9 to 12 months (Mühlau, Kaliszewska and Röder 2011; Platt, Luthra and Frere-Smith 2015), the four months recruitment period in our study was relatively short. This is also likely to have contributed towards the limited success of the approach. A longer recruitment period could have led to longer recruitment waves, possibly resulting in more diverse recruitment trees than observed in our study. Of particular interest would have been to observe whether, or at which stage recruitment between men and women would have become stable.

To speed up response in the afore-mentioned study by Frere-Smith, Luthra and Platt (2014), researchers asked seeds for contact details of the eligible persons acquainted with them and consequently contacted such persons themselves. In contrast we followed the standard RDS procedure, whereby contact to potential recruits is made at their own initiation (Heckathorn 1997). While adapting the procedure might have led to increased participation and faster recruitment, we believe that the request to divulge their acquaintances' contact details without the concerned person's acquiescence would have been met with resistance. Indeed, Frere-Smith and colleagues do report that some of their participants were less willing to give the requested contact details than others. Most importantly, data protection laws in Germany would not permit such an approach.

The fact that we did not follow-up on the seeds is a further limitation as we lack information on whether or not the coupons were distributed. Reminder calls might have encouraged or reminded them to distribute coupons and also helped shed light into factors that apparently made it difficult for male seeds to recruit a further person. We could then have used this information to make necessary adaptations where possible. Further, post-survey focus-group discussions with seeds could have helped provide insight into factors that need to be taken into consideration or improved on in future studies.

A further reason that could have contributed towards the limited success of the approach is the fact that recruited persons were required to call the respective study centre to make an appointment for the interview. The telephones were manned during set times by staff who spoke only German and English, thereby posing a barrier to those who could not call during

these times and those not proficient enough in either language. This problem however does not only apply to the RDS sample, but also to the RBS, as all prospective participants were required to make an appointment by telephone. Using an app based appointment system could possibly have benefitted the survey. Setting up the system however requires time, technological know-how and financial resources. In addition, persons not having smart phones or lacking sufficient computer competencies would have been excluded from the survey. We did not conduct an on-line survey for similar reasons. Offering to conduct the interviews at the participant's home rather than just site-specific, might however have improved survey participation in general. A combination of home-based interviews with flexible interview times that include early evening hours or possibly even Saturdays might have enabled more men to participate. The interview times offered, daytime on normal weekdays, were more suitable for persons not employed or with part-time employment, generally women, thereby possibly partly leading to the oversampling of women in both RBS and RDS. As has already been mentioned, the shortness of the recruitment waves, coupled with the fact that women were more successful seeds than men, also led to women being oversampled.

The difference in the way sampling progressed in the two neighbourhoods is also notable. While the fact that RBS was more successful in the wealthier neighbourhood, Neustadt than in Groepelingen is in line with findings that persons with a lower socio-economic profile participate less in health surveys (Demarest et al. 2013), we had expected that RDS would function better in the more impoverished neighbourhood, surmising that the offer of double incentives would be a great motivator. We offered 10 Euro for the own interview and a further 10 Euro for each recruited person, meaning that an individual could get up to 40 Euro. Although the 10 Euro was slightly above the then minimum wage of 8,84 Euro, it was nevertheless possibly perceived as being too little. Further, for some people contacting the study centre and going there for the interview probably meant them incurring costs, financial as well as time, which they perhaps felt were not reimbursed enough, particularly the time aspect.

The fact that compared to Neustadt, Groepelingen has in recent years received more different groups of new migrants as the rentals there are lower, possibly led to the very short recruitment trees and more distinct seed dependent groupings observed. Although RDS has been effectively used to recruit specific groups of migrants (Tyldum and Johnston 2014), Frere-Smith, Luthra and Platt (2014), argue that it might not necessarily be the most suitable way to reach non-specific migrant groups in diverse urban settings. Selecting the 'right' seeds in such neighbourhoods is also challenging due to the diversity and clustering of subpopulations as well as limited networks, which hinder networking at a larger level. One solution could be to recruit a large number of seeds as was done in the study by Mühlau, Kaliszewska and Röder (2011), in which 55 seeds were recruited. This, however, might induce a stronger bias due to seed dependence.

Conclusion

RDS can be a feasible, albeit challenging approach to recruit persons not normally reached using standard recruitment methods. To enable the optimisation of the approach, enough time and resources should be set aside for getting to know the characteristics of the study population of interest and selection of appropriate seeds. Further, mechanisms to follow up and support the seeds should be set up. The use of modern technology such as app-based appointment systems or online surveys should also be considered. Care should however be taken not to exclude those with limited or no access to the internet, respectively with limited or no computer competencies.

Compliance with ethical standards

Funding

Ethical approval
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent
Informed consent was obtained from all individual participants included in the study.

Conflict of interest
All authors declare no conflict of interest

References

Demarest S, Van der Heyden J, Charafeddine R, Tafforeau J, Van Oyen H, Van Hal G (2013) Socio-economic differences in participation of households in a Belgian national health survey. Eur J Public Health 23(6): 981-5

European Social Survey (2012). ESS Round 6 Source Questionnaire. London, Centre for Comparative Social Surveys, City University London

Frere-Smith T, Luthra R, Platt L (2014) Sampling recently arrived immigrants in the UK: Exploring the effectiveness of respondent driven sampling. CreAM Discussion Paper Series. London, UK, Center for Research and Analysis of Migration. CPS 32/14

George S, Duran N, Norris K (2014) A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 104(2): e16-31

Gile KJ (2011) Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. J Am Stat Assoc 135-146

Górny A, Napierała J (2016) Comparing the effectiveness of respondent-driven sampling and quota sampling in migration research. Int J Soc Res Methodol 19(6): 645-661

Handcock MS, Fellows IE, Gile KJ (2014) RDS Analyst: Software for the analysis of respondent-driven sampling data, Version 0.42. http://hpmrg.org. Accessed January 2018

Heckathorn D (1997) Respondent-driven sampling: a new approach to the study of hidden populations Social Problems 44(Suppl 2): 174-199

Hu J, Wong KC, Wang Z (2015) Recruiting migrants for health research through social network sites: an online survey among Chinese migrants in Australia. JMIR Res Protoc 4(2): e46

Johnston LG, Thurman TR, Mock N, Nano L, 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

Johnston LG (2014) Sampling migrants: How respondent driven sampling works. In Tydum G and and Johnston LG (ed) Applying respondent driven sampling to migrant populations: Lessons from the field. Palgrave Macmillan, Basingstoke, UK pp 9-16

Johnston LG, Luthra R (2014) Analyzing data in RDS. In: Tydum G and Johnston LG (ed) Applying respondent driven sampling to migrant populations: Lessons from the field. Palgrave Macmillan, Basingstoke, UK, pp 84-99

Johnston LG, Luthra,L, Napierala J (2014) Measuring personal network size in RDS. In: Tydum G and Johnston LG (ed) Applying respondent driven sampling to migrant populations: Lessons from the field. Palgrave Macmillan, Basingstoke, UK, pp 27-36

Johnston LG, Hakim AJ, Dittrich S, Burnett J, Kim E, White RG (2016) A systematic review of published respondent-driven sampling surveys collecting behavioral and biologic data. AIDS Behav 20(8): 1754-1776

Kubal A, Shvab I, Wojtynska A (2014) Initiation of the RDS recruitment process: Seed selection and role. In Tyldum G and Johnston LG (ed) Applying Respondent Driven Sampling to Migrant Populations: Lessons from the field. Palgrave Macmillan: Basingstoke, UK pp 37-47

Montealegre JR, Risser JM, Selwyn BJ, McCurdy SA, Sabin K (2013) Effectiveness of respondent driven sampling to recruit undocumented Central American immigrant women in Houston, Texas for an HIV behavioral survey. AIDS Behav 17(2): 719-727

Mühlau P, Kaliszewska M, Röder A (2011) Polonia in Dublin: Preliminary Report of Survey Findings. Dublin: Trinity College Dublin. https://www.tcd.ie/sociology/assets/pdf/Polonia in Dublin Report I[1].pdf. Accessed October 2018

OECD (2011) Divided We Stand - Why inequality keeps rising. doi:http://dx.doi.org/10.1787/9789264119536-en

Pelikan JM, Röthlin F, Ganahl K (2014) Measuring comprehensive health literacy in general populations: validation of instrument, indices and scales of the HLS-EU study. 6th Annual Health Literacy Research Conference. Bethesda, Maryland, USA

Phillimore J, Bradby H, Knecht M, Padilla B, Brand T, Cheung SY, Pemberton S, Zeeb H (2015) Understanding healthcare practices in superdiverse neighbourhoods and developing the concept of welfare bricolage: Protocol of a cross-national mixed-methods study. BMC Int Health Hum Rights 15: 16

Platt L, Luthra L, Frere-Smith T (2015) Adapting chain referral methods to sample new migrants: possibilities and limitations. Demographic Res. 33: 665-700

Quay TA, Frimer L, Janssen PA, Lamers Y (2017) Barriers and facilitators to recruitment of South Asians to health research: a scoping review. BMJ Open 7(5): e014889

Redwood S, Gill PS (2013) Under-representation of minority ethnic groups in research--call for action. Br J Gen Pract 63(612): 342-343

Reineke A, Pigeot I, Ahrens W, Rach S (2018) MODYS - A modular control and documentation system for epidemiological studies. in Bammann K, Lissner L, Pigeot I, Ahrens W (ed), Instruments for health surveys in children and adoloscents. Springer Nature: Switzerland, pp 25-45

Reiss K, Dragano N, Ellert U, Fricke J, Greiser KH, Keil T, Krist L, Moebus S, Pundt N, Schlaud M, Yesil-Jurgens R, Zeeb H, Zimmermann H, Razum O, Jockel KH, Becher H (2014) Comparing sampling strategies to recruit migrants for an epidemiological study. Results from a German feasibility study. Eur J Public Health 24(5): 721-726

Salganik M J (2006) Variance estimation, design effects, and sample size calculations for respondent-driven sampling. J Urban Health 83(6 Suppl): i98-112

Senatorin für Arbeit, Frauen, Gesundheit, Jugend und Soziales (2009) Lebenslagen in Bremen: Armuts- und Reichtumsbericht für das Land Bremen 2009. Bremen

Shaghaghi A, Bhopal RS, Sheikh A (2011) Approaches to recruiting 'hard-to-reach' populations into reearch: a review of the literature. Health Promot Perspect 1(2): 86-94

Statistisches Landesamt Bremen (2015) Bremen Kleinräumig Infosystem. https://www.statistik.bremen.de/ Accessed January 2018

Statistisches Landesamt Bremen (2016) Bremen Kleinräumig Infosystem. https://www.statistik.bremen.de/ Accessed January 2018

Strathdee SA, Lozada R, Ojeda VD, Pollini RA, Brouwer KC, Vera A, Cornelius W, Nguyen L, Magis-Rodriguez C, Patterson TL, Proyecto El C (2008) Differential effects of migration and deportation on HIV infection among male and female injection drug users in Tijuana, Mexico. PLoS One 3(7): e2690

Tyldum G, Johnston LG (Eds) (2014) Applying respondent driven sampling to migrant populations: Lessons from the field. Basingstoke, UK, Palgrave Macmillan

United Nations (2016). International Migration Report 2015 - Highlights. United Nations. ST/ESA/SER.A/375

Vertovec S (2007) Super-diversity and its implications. Ethnic and Racial Studies 30(6): 1024-1054

Ware JE, Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30(6): 473-483

Weber MF, Banks E, Smith DP, O'Connell D, Sitas F (2009) Cancer screening among migrants in an Australian cohort; cross-sectional analyses from the 45 and Up Study. BMC Public Health 9: 144

Table 1: Comparison of register-based sample characteristics (n=779) with respondent-driven sampling based population prevalence estimates (Gile's SS) calculated based on the respondent-driven sampling sample (n=115) and available underlying population size figures for the two neighbourhoods (69 259 for 2015 and 70 118 for 2016). (Understanding the practice and developing the concept of welfare bricolage [UPWEB Study], Bremen, Germany, 2017)

	RBS sample (%)	Actual population size (%)	RDS-based estimate (%)	95% Confidence Interval	Estimated Design Effect
Sex					
Female	57.0	49.0	67.2	54.8 - 79.5	2.08
Male	43.0	51.0	32.8	20.5 – 45.2	
Age					
≤45	36.1	47.4	48.8	32.6 – 64.9	3.15
46-60	27.3	24.0	23.1	12.4 – 33.9	1.94
>60	36.6	28.6	28.1	14.4 – 41.9	2.81
Migrant background					
Yes	32.5	34.8	51.6	40.4 – 62.7	1.51
No	67.5	65.2	48.5	37.3 – 59.7	
Unemployed					
Yes	7.5	19.6	18.1	8.4 –27.9	1.93
No	92.5	80.4	81.9	72.1 – 91.6	
¤Poverty risk					
Yes	30.4	-	55.0	41.3 – 68.7	2.24
No	69.6		45.0	31.3 – 58.7	
¤Educational level[*2]	14.8		23.9	11.6 – 36.1	2.47
Low	45.7	-	36.8	26.8 – 46.8	1.29
Medium	39.5		39.3	25.0 – 53.7	2.59
High					
¤Interview language		-			
German	96.5		87.3	74.1 – 100.6	4.74
Other	3.5		12.7	0.0 – 25.9	
¤German language proficiency[§]					
Poor/average	24.6	-	43.7	22.3 – 65.0	1.88
Good/very good	75.4		56.4	35.0 – 77.7	
¤Discrimination[*5]					
Yes	8.0	-	13.9	6.6 – 21.2	1.31
No	92.0		86.1	78.8 – 93.4	
¤Self-reported health[*1]					
Fair/poor	16.6		21.0	11.1 – 30.8	1.75
Good	49.0	-	33.4	23.6 – 43.3	1.29

Excellent/very good	34.4		45.6	32.8 – 58.4	1.97
¤**Health literacy**					
Low	9.0		14.1	6.9 – 21.4	1.30
Medium	74.8	-	68.8	59.1 – 78.4	1.30
High	16.2		17.1	9.0 – 25.2	1.39

RBS – register-based sampling; RDS – respondent-driven sampling; ¤No underlying population data available; *x Missing data for x participants (e.g. Education level*2); §Applies only for those not born in Germany, missing data for 38 participants, leaving n=205 for analysis

Fig1. Recruitment tree for respondent-driven sampling with nodes differentiated by neighbourhood and sex. (Understanding the practice and developing the concept of welfare bricolage [UPWEB Study], Bremen, Germany, 2017)
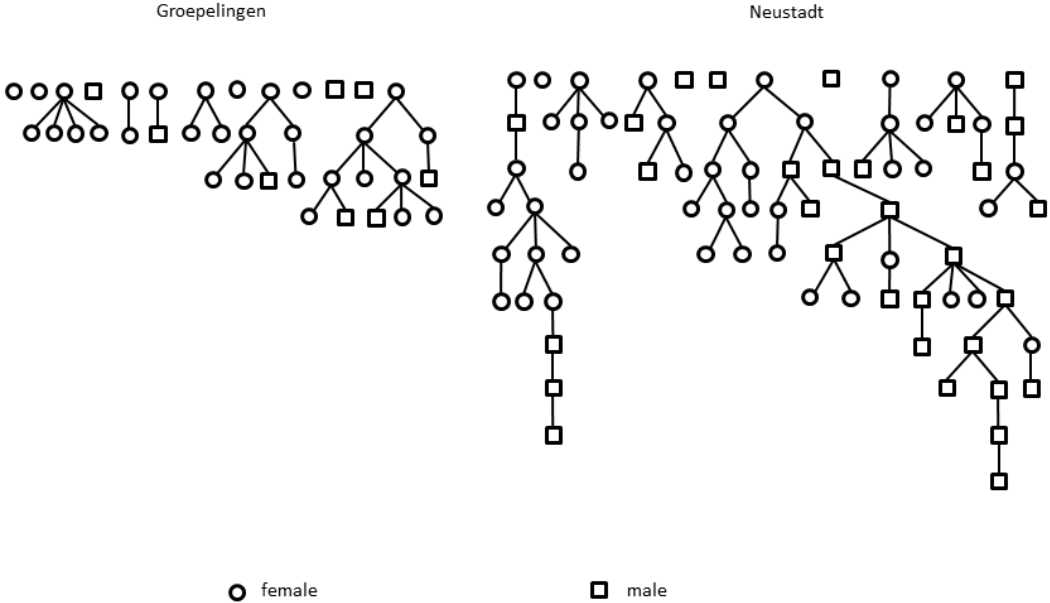
Fig 2. Recruitment tree for respondent-driven sampling for the two neighbourhoods, Groepelingen and Neustadt, with nodes differentiated by migrant background and poverty risk. (Understanding the practice and developing the concept of welfare bricolage [UPWEB Study], Bremen, Germany, 2017)