Using Respondent Driven Sampling to Measure the Incidence and Prevalence of Informal Sector Abortion: A Methodological Assessment

Ruvani T. Jayaweera[1,2], M.G. Isabel Muñoz[1,2], Relebohile Motana[1], Tshego Bessenaar[1], Caitlin Gerdts[1]

1. Ibis Reproductive Health
2. University of California, Berkeley, School of Public Health, Department of Epidemiology

ABSTRACT

Unsafe abortion is a preventable cause of maternal mortality and morbidity, yet data on the prevalence and effects of unsafe abortion are fraught with bias. Respondent Driven Sampling (RDS) is a sampling methodology that has been successfully used to estimate the prevalence of sensitive and illegal behaviors among hidden populations. This paper explores the feasibility of applying the RDS methodology to estimate the incidence and lifetime prevalence of informal sector abortion.

The validity of RDS relies on several key assumptions: the population being recruited must know each other as members of the target population and form social ties based on this characteristic, the referral process should result in a series of overlapping networks, and sampling should replicate sampling with replacement. To test the assumptions of using RDS to measure informal sector abortion incidence and prevalence, we are conducting a study in Soweto, South Africa. Success of this method will be determined by generating an RDS sample that has similar demographics to the overall population of women of reproductive age living in Soweto. An advantage of this study is that demographics of the target population are known from census data, unlike most other applications of RDS for which no established sampling frame for the population of interest exists. Other metrics evaluated will be time to equilibrium, length of referral chains, reciprocity of network ties, and random recruitment within the recruiter's personal network.

Results from this study can be used to evaluate the feasibility and appropriateness of using RDS to effectively recruit participants and measure informal sector abortion among a population where such abortions are common; positive results could demonstrate that RDS is an exciting new tool for producing unbiased estimates of prevalence, safety, efficacy, and experiences with informal sector and/or unsafe abortion in various contexts.

INTRODUCTION

Unsafe abortion is a completely preventable cause of maternal mortality and morbidity worldwide[1], yet is responsible for an estimated 8-13% of all maternal deaths worldwide.[2] In settings where abortion is legally restricted, or where it is permitted but not widely accessible, women face significant barriers to high quality abortion care. In some settings, women are increasingly choosing mifepristone and/or misoprostol to terminate their pregnancies on their own[3], or seeking abortion from clandestine providers in potentially unsafe conditions[4]. There is a pressing need for new and innovative research methods that can accurately measure the incidence and characteristics of abortions that occur outside of the formal health care setting, referred to in this paper *as informal sector abortions*[5].

Existing data on the prevalence and effects of informal sector abortion are fraught with bias. Indirect methods for estimating the prevalence of induced abortions rely on assumptions and extrapolations that are often difficult to test[6,7], and direct survey techniques suffer from widespread underreporting[7-9]. Due to social and cultural stigma, and fear of legal consequences, women seeking informal sector abortion may be reluctant to seek care in the event of complications, and may also be reluctant to reveal their (often illegal) abortion experience to researchers [10-13]. While recent evidence suggests that over half of all abortions that occur globally are performed in illegal or unsafe conditions[6], challenges in data collection likely lead to underestimates of abortion prevalence, as well as biased data on the characteristics and outcomes of abortion in such contexts. Accurately measuring the incidence of abortion, and understanding the scope of informal sector abortions in settings where legal abortion exists can document the gaps in accessibility of these services and help identify interventions to address issues around access to and quality of care.

Overview of Respondent Driven Sampling

Due to the stigma and secrecy associated with informal sector abortion and abortion in general, new data collection and estimation strategies are needed. Respondent Driven Sampling (RDS) is a sampling methodology that has been successfully used to estimate unbiased prevalence of sensitive and illegal behaviors among hidden populations such as injection drug users, sex workers, and men who have sex with men[7-13].

RDS relies upon social networks to identify populations engaging in stigmatized, illicit, or otherwise hidden behaviors; this approach may offer a previously untested alternative to measuring informal sector abortions. RDS attempts to leverage a small non-random sample of initial

participants (known as *seeds*) within social networks engaging in hidden or stigmatized behaviors to recruit others within the same social networks (i.e. the target population). Each individual seed is given a set number of coupons with which they can recruit their social network peers. Once a participant with a valid coupon presents to the study site, she is provided with the same number of coupons with which to enroll other members of the social network, thus resulting in a lengthy chain of participants representing the target population[9 10 14]

In order to generate unbiased prevalence estimates for the outcomes, behaviors, or characteristics of interest in the target population, RDS samples are adjusted for potential selection bias in analyses by weighting participants with more contacts in the target population inversely proportional to the number of contacts in the network itself [9 10 15]. Like any statistical methodology, RDS has its limitations—specifically that accuracy of estimates can be influenced by recruitment dynamics and the distribution of the behavior of interest within the social network [15] —however, recent research suggests that rigorous formative research to identify appropriate initial seeds can improve the accuracy of estimates[7]. Testing the feasibility of using RDS to effectively recruit participants and measure informal sector abortion among a population where such abortions are common could result in an exciting new tool for producing unbiased estimates of prevalence, safety, efficacy, and experiences with informal sector and/or unsafe abortion in various contexts.

Application of RDS to the study of informal sector abortion

While RDS was initially developed to identify "hidden" populations engaging in stigmatized, illicit, or otherwise hidden behaviors, it may offer a previously untested alternative to measuring informal sector abortions. We posit that women of reproductive age who live in settings where access to abortion is restricted are "hidden" populations, as they are missed by traditional sampling methods utilized in most studies of informal sector abortion incidence. For example, facility based studies may miss women who do not seek care, who may be systematically different in their experiences with informal sector abortion than those who present for services at clinics where the study is being conducted. RDS may be a more efficient approach to accessing a broader, more representative sample of women in which a less biased estimate of informal sector abortion incidence can be calculated. Ample evidence suggests that population-based household survey estimates of experiences with abortion, when asked directly, underestimate the true incidence of abortion[24]; this bias is likely further amplified in settings where abortion is highly stigmatized and/or criminalized. Another advantage of RDS is the possibility that the process of being recruited to the

study by a member of one's peer network engenders trust between the participant and the interviewer; thus, RDS may potentially decrease the likelihood of underreporting of abortion.

The validity of RDS as a sampling methodology relies on several key assumptions[15,16,23]. First, the population being recruited must know each other as members of the target population, and form social ties on the basis of this shared characteristic. Second, the referral process should result in a series of overlapping networks (networks of networks), rather than isolated referral chains. Third, sampling should replicate sampling with replacement. Additional assumptions are that participants can accurately report their network size and are randomly sampling their recruits from within this personal network. Studies employing RDS to study informal sector abortion should aim to evaluate these assumptions.

In order to generate a less-biased, population-representative estimate of the lifetime prevalence of informal sector abortion, as well as test the assumptions underlying this novel application of the RDS method, we are conducting a Respondent Driven Sampling study in Soweto, South Africa. This paper presents our study methodology and explores the feasibility of applying RDS methodology to estimate the prevalence of informal sector abortion.

FIELD TESTING THE RDS METHOD

Study setting

The study is being conducted in Soweto, South Africa, a large township in the City of Johannesburg. Johannesburg is the largest city in South Africa, with a population of approximately 5 million people. Women of reproductive age (15-49 years old) represent 30.2% of the total population, and 61.1% of the total female population in Johannesburg[16]. In South Africa an estimated 50% of abortions take place outside designated health facilities[17]; and according to the WHO 90% of self-induced abortions in South Africa are unsafe[18]. Given that abortion is a right protected by the Choice of Termination Pregnancy Act, reliable information about the prevalence of abortions outside of the formal health system is needed [19]. Ethical approval for this study was granted by the Human Sciences Research Council in Pretoria, South Africa.

Study participants and recruitment

We established relationships with a variety of community-based organizations in Soweto in order to identify a diverse range of women of reproductive age to serve as "seeds" to begin RDS

recruitment. While in RDS methodology, the final sample is ultimately independent of the composition of the initial seeds, diversity of seeds may decrease the time until the sample reaches equilibrium, equilibrium, or the point of recruitment at which the demographic composition of the study sample does not change drastically with each successive wave of recruitment.[14 20]

Seeds will respond to an initial interviewer-administered questionnaire regarding their experiences with abortion as well as information about the composition of their social network. After completing this questionnaire, seeds will be given three coupons with which to recruit other eligible participants, known as *recruits*, who are members of their social network. Eligibility criteria includes being a woman between the ages 15-49; able to speak English, Tswana, isiZulu, Sotho, or Xhosa; and living in Soweto. Parental consent will be obtained from participants who are under the age of 18.

Recruits are instructed to call the study phone number listed on their recruitment coupon; project staff will assess eligibility and schedule an interview at a time and study site of the participant's choosing. At the interview site, eligible participants who consent to participate in the study will respond to the interview-administered questionnaire. After completing the survey, these participants will also receive three coupons with which to recruit additional participants from their social network. Participants will receive an incentive of R75 (~ USD 6) for completing the main RDS questionnaire, and a secondary reimbursement of R50 (~ USD 4) for each successfully recruited study participant. Tracking referral chains is essential to adjusting for potential selection bias as a result of seed dependence[14]. Each coupon contains a unique identifier which will allow the study team to track recruits and link them to their corresponding recruiters.

When participants return to claim their secondary reimbursement for study participants whom they successfully recruited, they will complete a brief follow-up survey about their experiences participating in and recruiting for the study. If a participant indicated a prior history of abortion (either in the informal or formal sector) in the initial survey, they will be invited to participate in an additional survey and receive an incentive of R75 (~ USD 6). Survey data will be collected using Qualtrics[21].

Interview sites in Soweto were chosen because they provided a private and quiet space, and were known to and easily accessible for the target population. Sites will remain open for at least one week after the last coupon is distributed. Study recruitment began in early April 2018 and is expected to continue through at least October 2018.

Sample size

Based on the best estimates of the prevalence of informal sector abortion in the general population, as well as RDS recruitment rates per seed in other stigmatized populations in South Africa[22 23], we aim to recruit a sample size of 900 women to participate in the study. Recruitment numbers and convergence towards selected sociodemographic characteristics will be closely monitored. No further data collection will continue one-year after recruitment has begun. After 900 participants have completed the main RDS survey, and equilibrium has been reached, no new coupons will be distributed.

Instruments

The main RDS instrument is a quantitative questionnaire with questions on demographic characteristics, social network composition, health behaviors, sexual history, pregnancy history, and experience with abortion. The RDS follow-up instrument has questions on the recruiting process, including questions on refusals. The abortion follow-up instrument has questions on participants' abortion experience, including questions on feelings of preparedness, cost, treatment from providers, as well as attitudes towards abortion and contraception decision-making. Categories for informal sector providers and methods were informed by previous literature[19 24 25], as well as findings from formative research for this study, in which in-depth interviews were conducted with 19 women from Soweto who had attempted to terminate a pregnancy outside of the formal health setting[26].

We conducted cognitive interviews with five participants who participated in the in-depth interviews to test key questions from the study instruments. Results from these interviews resulted in the refinement of key questions; for example, "terminating a pregnancy" and "abortion" was preferred to the term "ending a pregnancy," which was easily confused with pregnancies that ended in stillbirth or miscarriage. Similarly, "backstreet abortion" was widely accepted as referring to abortions in the informal sector, while "TOP" (which stands for termination of pregnancy) was understood as abortions that took place in a formal health setting. The term "induced abortion" was not understood by participants and is not used in our final instruments.

The primary outcome of interest for this study is lifetime prevalence of informal sector abortion, measured as the proportion of women in the study who report at least one informal abortion in their lifetime, adjusted for the RDS sampling methodology. The secondary outcome of interest for this study is 1-year and 5-year incidence of abortion (both in the informal sector and in the formal health setting). Key sociodemographic variables to compare representativeness with the

source population are age, educational attainment, employment status, and home language. Essential for calculating the RDS estimator is network degree, which is assessed by the following question: "How many women of reproductive age who live in Soweto do you know, who also know you, that you have seen in the past week?" This measure of network degree is in accordance with other RDS studies[27 28].

Analysis

Data management will be conducted in R[29] and Stata[30]. The estimated proportion of the population that has ever had an informal sector abortion will be calculated using the RDS-2 estimator[31] in R. This proportion is calculated as the number of respondents who report ever having had an informal sector abortion, weighted by the inverse of their network degree size (Equation 1).

$$\hat{p} = \frac{\sum_{j \in I} \frac{1}{d_j}}{\sum_{j \in S} \frac{1}{d_j}}$$

Equation 1: RDS-2 Estimator; where *j* indexes the respondent, ***S*** is the set of the full sample, ***I*** is the set of respondents who have ever received an informal sector abortion, d_j is the self-reported network degree.

Testing RDS assumptions

In order to assess the feasibility of applying RDS methodology to estimate the prevalence of informal sector abortion in Soweto, South Africa, the core assumptions[32] underlying RDS should be rigorously assessed. Assumptions and proposed methods for assessing these assumptions are reviewed below.

1) Respondents are able to accurately report their degree within a network.

While the true size of a respondent's degree (network size) cannot be objectively verified, we will assess the distribution of reported degree size to assess plausibility, as well as for evidence of rounding. As mentioned above, our measure of degree is ascertained by the question, "How many women who live in Soweto and are of reproductive age (15-49) do you know by name and they know you by name?" and "Of these women, how many have you seen in the past week?" Respondents are probed to think through categories of individuals (i.e. family members, friends, co-workers, neighbors) in order to minimize the tendency to estimate or round. At follow-up,

participants will be asked to report their degree again; this will be validated against their reported degree in the main instrument, and the reliability of this measure will be used to assess this assumption.

2) Respondents randomly recruit from their personal network.

Respondents are asked, "Of these women (in your social network, who you have seen in the past week), how many are currently working?". The overall proportion of those reported to be currently working from possible recruits (degree size) will be compared to the proportion of those working in the study sample, as a crude assessment of random recruitment. As we have reasonable estimates of demographics of the target population from census data, we will also be able to compare the composition of our study population to the target population. In addition, respondents are asked at follow-up whether anyone they approached declined to accept a coupon; this will be used to calculate refusal rates. Lastly, respondents are asked an open-ended question "How did you decide who to give the coupon to?"; responses to this question will be coded thematically to assess the relationship between this decision and random recruitment.

3) Network connections are reciprocated.

First, the population being recruited must know each other as members of the target population, and form social ties on the basis of this shared characteristic. In an RDS study of informal sector abortion among women of reproductive age, the target population of interest is women of reproductive age. As women likely form social ties with other women based on their shared group membership, it is highly likely that an RDS study using this eligibility criteria meets this assumption. In order to assess whether these ties are reciprocated, respondents are asked "How do you know the person who recruited you?" at their main interview; at follow-up, respondents are also asked "How do you know the person who you recruited?" The proportion of respondents who report that they did not have a personal relationship with the person who recruited them will serve as an indication of whether this assumption has been met. In addition, participants are asked, "Do you think that the person to whom you gave a coupon to would have given you a coupon if you had not participated in the study?", which serves as an additional check on the reciprocity assumption.

4) The sample composition converges to equilibrium.

After sufficient waves, the composition of the overall sample should be independent of the initial seeds, in order to overcome the non-randomness of how the seeds were recruited[32]. Sampling should replicate sampling with replacement; as women of reproductive age are a large proportion of the total population, even an RDS study employed in relatively small geographical area is unlikely to have a large sampling fraction. Thus, the sample size recruited by RDS will likely be small relative to the overall size of the target population, mimicking a sampling with replacement process. Time to convergence of key sociodemographic characteristics, as well as the overall percentage of the population that reported ever having an informal sector abortion, will be monitored. In addition, unlike other RDS studies, there is existing sociodemographic data on the target population that can be used to compare the overall representativeness of the sample to the underlying population.

An additional proposed benefit of RDS is the possibility that the process of being recruited to the study by a member of your peer network engenders trust between the participant and the interviewer; thus, RDS may potentially decrease the likelihood of underreporting of abortion. To test this assumption, we have included questions about other potentially stigmatized and underreported health behaviors, tobacco and alcohol use in the main instrument. Increased reporting of these behaviors, relative to what is reported in the current South Africa Demographic and Health Survey may serve as an indication that the recruitment method encourages disclosure of under-reported health behaviors.

CONCLUSION

While Respondent Driven Sampling holds great promise for generating less biased, population-based estimates of informal sector abortion behaviors, to our knowledge, RDS has never been employed for generating informal sector abortion incidence or prevalence estimates. Accurate estimates are vital to developing targeted and effective programs, policies, and interventions to increase access to safe abortion and to improve women's health. We explored the feasibility of applying the RDS methodology to estimate the incidence and lifetime prevalence of informal sector abortion, and outlined key assumptions that should be rigorously tested to demonstrate the potential of future applications of this method.

Success of the RDS method will be determined by generating a study sample that has similar demographics to the overall population of women of reproductive age living in Soweto, as well as an assessment of whether the RDS assumptions have been met. An advantage of this study is that

demographics of the target population are known from census data, unlike most other applications of RDS for which no established sampling frame for the population of interest exists. While McCreesh et al found that RDS produced a generally representative sample in a non-hidden population, issues with bias were still present, mainly due to failure to meet certain RDS assumptions[28]. This study will attempt to quantifiably assess the RDS assumptions in order to determine the validity of our estimate. Metrics evaluated will be time to equilibrium, or the point of recruitment at which the demographic composition of the study sample does not change drastically with each successive wave of recruitment, as well as the length of referral chains, and reciprocity of network ties. Other assumptions key to the validity of the RDS estimate, such as random recruitment within the recruiter's personal network, will be directly evaluated through items in the follow-up questionnaire.

The potential advantages of RDS over existing methods for the measurement of informal sector abortion far outweigh the potential limitations. Results from this study can be used to evaluate the feasibility and appropriateness of using RDS to effectively recruit participants and measure informal sector abortion among a population where such abortions are common; positive results could demonstrate that RDS is an exciting new tool for producing unbiased estimates of prevalence, safety, efficacy, and experiences with informal sector and/or unsafe abortion in various contexts. RDS may have broad-reaching implications for abortion research globally as a mechanism to produce less-biased estimates of key abortion outcomes.

REFERENCES

1. Grimes DA, Benson J, Singh S, et al. Unsafe abortion: the preventable pandemic. *The Lancet* 2006;368(9550):1908-19.
2. Ganatra B, Gerdts C, Rossier C, et al. Global, regional, and subregional classification of abortions by safety, 2010-14: estimates from a Bayesian hierarchical model. *Lancet* 2017;390(10110):2372-81. doi: 10.1016/S0140-6736(17)31794-4 [published Online First: 2017/10/02]
3. Hyman A, Blanchard K, Coeytaux F, et al. Misoprostol in women's hands: a harm reduction strategy for unsafe abortion☆. *Contraception* 2013;87(12):8-130.
4. Grossman D, Holt K, Peña M, et al. Self-induction of abortion among women in the United States. *Reproductive health matters* 2010;18(36):136-46.
5. Sedgh G FV, Owolabi OO, Singh SD, Askew I, Bankole A, Benson J, Rossier C, Pembe AB, Adewole I, Ganatra B. . Insights from an expert group meeting on the definition and measurement of unsafe abortion. *International Journal of Gynecology & Obstetrics* 2016 doi: 10.1016/j.ijgo.2015.11.017
6. Sedgh G, Singh S, Shah IH, et al. Induced abortion: incidence and trends worldwide from 1995 to 2008. *The Lancet* 2012
7. Wang J, Carlson RG, Falck RS, et al. Respondent-driven sampling to recruit MDMA users: a methodological assessment. *Drug and alcohol dependence* 2005;78(2):147-57.
8. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Social problems* 1997:174-99.
9. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Social problems* 2002;49(1):11-34.
10. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological methodology* 2004;34(1):193-240.
11. Decker MR, Marshall BD, Emerson M, et al. Respondent-Driven Sampling for an Adolescent Health Study in Vulnerable Urban Settings: A Multi-Country Study. *Journal of Adolescent Health* 2014;55(6):S6-S12.
12. Reisner SL, Mimiaga MJ, Johnson CV, et al. What makes a respondent-driven sampling "seed" productive? Example of finding at-risk Massachusetts men who have sex with men. *J Urban Health* 2010;87(3):467-79.

13. Ludlam AH, Saxton PJ, Dickson NP, et al. Respondent-driven sampling among gay and bisexual men: experiences from a New Zealand pilot study. *BMC research notes* 2015;8(1):549.
14. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl* 1997;44(2):174-99.
15. Goel S, Salganik MJ. Assessing respondent-driven sampling. *Proceedings of the National Academy of Sciences* 2010;107(15):6743-47.
16. Africa SS. City of Johannesburg [Available from: http://www.statssa.gov.za/?page_id=1021&id=city-of-johannesburg-municipality.
17. HEARD. Unsafe abortion in South Africa country fact sheet: Durban: Health Economics and HIV/AIDS Research Division / University of KwaZulu-Natal 2016.
18. World Health Organization. Unsafe abortion: global and regional estimates of incidence of unsafe abortion and associated mortality in 2008. 2011
19. Gerdts C, Raifman S, Daskilewicz K, et al. Women's experiences seeking informal sector abortion services in Cape Town, South Africa: a descriptive study. *BMC Womens Health* 2017;17(1):95. doi: 10.1186/s12905-017-0443-6 [published Online First: 2017/10/04]
20. Gile KJ, Johnston LG, Salganik MJ. Diagnostics for Respondent-driven Sampling. *J R Stat Soc Ser A Stat Soc* 2015;178(1):241-69. doi: 10.1111/rssa.12059 [published Online First: 2014/05/01]
21. Qualtrics I. Qualtrics. com. *Provo, UT, USA* 2013
22. Lane T, Raymond HF, Dladla S, et al. High HIV prevalence among men who have sex with men in Soweto, South Africa: results from the Soweto Men's Study. *AIDS and Behavior* 2011;15(3):626-34.
23. Townsend L, Jewkes R, Mathews C, et al. HIV risk behaviours and their relationship to intimate partner violence (IPV) among men who have multiple female sexual partners in Cape Town, South Africa. *AIDS and Behavior* 2011;15(1):132-41.
24. Constant D, Grossman D, Lince N, et al. Self-induction of abortion among women accessing second-trimester abortion services in the public sector, Western Cape Province, South Africa: an exploratory study. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* 2014;104(4):302-5. [published Online First: 2014/08/15]
25. Harries J, Gerdts C, Momberg M, et al. An exploratory study of what happens to women who are denied abortions in Cape Town, South Africa. *Reprod Health* 2015;12:21. doi: 10.1186/s12978-015-0014-y [published Online First: 2015/04/18]

26. Ramirez A, Jayaweera R, Mtimkulu M, et al. Understanding abortion seeking and care outside of legal public sector services in South Africa: a qualitative study. *Contraception* 2017;96(4):277-78.
27. McCormick TH, Salganik MJ, Zheng T. How many people do you know?: Efficiently estimating personal network size. *Journal of the American Statistical Association* 2010;105(489):59-70.
28. McCreesh N, Frost S, Seeley J, et al. Evaluation of respondent-driven sampling. *Epidemiology (Cambridge, Mass)* 2012;23(1):138.
29. Team RC. R: A language and environment for statistical computing. 2013
30. LP SC. Stata Statistical Software Release 15: Stata Press Publication 2017.
31. Salganik MJ. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health* 2006;83(1):98.
32. Volz E, Heckathorn DD. Probability based estimation theory for respondent driven sampling. *Journal of official statistics* 2008;24(1):79.