

# Surveying Rare or Hidden Populations Using a Probability-Based Household Panel

Vicki J Pineau, J Michael Dennis,  
Stuart Michaels, Sherry Emery, Nada Ganesh  
NORC at the University of Chicago

## Abstract

Estimating economic, health, and social disparities among priority subpopulations (e.g., sexual and gender minorities) is increasingly regarded as essential for policy making and scientific inquiry but is problematic without resorting to non-probability sampling. Traditional probability-based sampling strategies are impractically expensive because of the large scale in-field screening required to find sufficient numbers of persons in so-called “rare” or “hidden” populations. As a result, rare or hidden populations are often studied using less rigorous methods such as “snow-ball” sampling and non-probability, opt-in web panels. Our paper is based on an NORC pilot study which tested a cost-effective alternative for surveys of rare or hidden populations. The tested approach combines probability sampling and Network and Respondent-Driven Sampling (RDS). The target subpopulations for the pilot study are lesbian, gay, bisexual, and transgender (LGBT) Americans. Pilot study findings are presented with respect to: (1) assessing the feasibility of using a probability-based panel sample, with multiple rounds of nominations of non-AmeriSpeak panelists by AmeriSpeak panelists, to survey a larger sample of people who self-identify as L, G, B, or T in the U.S; and (2) comparing estimates of LGBT status and health outcome estimates between the seed sample and the final sample. We also consider the viability of two alternative estimators for the pilot sample – an estimator using a multiplicity counting approach synonymous with Network Sampling, and an estimator under an approach based on RDS. The initial sample source for the pilot study is NORC’s AmeriSpeak Panel®, which is a household, multi-client panel that uses the probability-based NORC National Frame to construct an address-based nationally representative sample panel with sample coverage of approximately 97% of US households.

**Key Words:** Network Sampling, Respondent Driven Sampling, Rare and Hidden Populations, LGBT

## 1. Introduction

Respondent driven sampling (Heckathorn, 1997) is an extension of Network sampling (Sirken, 1998), whereby a chain of households or persons from the original sampled household is created organically from family and social networks through a referral protocol. The newly nominated households/persons are also interviewed in the course of fielding. Respondent driven sampling, or RDS, does not require a probability based sample as a starting point, relying on social network theory to calculate RDS weights and survey estimates for the population of interest. It differs from Network sampling in that it does not rely on a probability based sample as a starting point and probabilities of selection are not computed for newly recruited members from the RDS referral protocol. RDS is regarded as a fit-for-purpose approach, used when standard probabilistic methods are not practical nor cost effective, or because the target population is hard-to-reach. RDS examples include sample surveys of men who have sex with men, and active injection drug users. (Schneider, et al., 2013) (Gallagher, Sullivan, Lansky, & Onorata, 2017) (Schneider, et al., 2017) Other examples include special populations of artists and musicians or other specialists who are not part of a trade association, agritourism farmers, and migrant workers.<sup>1</sup> Typically, RDS studies are carried out at venues that are selected purposively and utilizing convenience/intercept samples with in-person, paper and pencil, or computer assisted interviewing modes.

A new methodology for conducting surveys of rare/hidden populations explored by NORC in a 2017 pilot survey includes aspects of Network sampling and RDS, seeking to take advantage of the strengths of both methodologies in conjunction with sample quality of the AmeriSpeak panel. In our pilot, we explored the use of RDS with the following enhancements:

1. Utilized a probability based sample from AmeriSpeak as a starting point for RDS referral methods;
2. Utilized the Web to generate the RDS referral sample;
3. Collected survey data on the multiplicities of respondents who were obtained through RDS referrals/recruitment under a sibling counting rule in order to investigate estimation under both the traditional Network Sampling approach and under the RDS approach.

Though RDS does not require a probability based sample as a starting point, use of a probability based sample from AmeriSpeak provided a less clustered and theoretically more random starting point than purposive or intercept sampling. The AmeriSpeak starting sample also provides a basis to assist in evaluating the representativeness of the RDS referred sample in terms of socio-demographic distributions by comparing the RDS referred sample to the starting sample.

---

<sup>1</sup> A compendium of articles and papers authored by experts in network and respondent driven sampling is available at: <http://www.respondentdrivensampling.org>.

The rare population of interest in our pilot is the LGBT population and our survey questions were about smoking behaviors and health. In recent years, sexual orientation and gender identity (SOGI) issues have become more salient in survey research. For example, a recent review of relevant research by the Kaiser Family Foundation found that segments of lesbian, gay, bisexual, and transgender populations can experience a wide range of disparities including inadequate access to health care services, earlier onset of disability and greater prevalence of mental health concerns. (Kates & Ranji, 2014) And a recent study of depression among the LGBT population reported far higher rates of depression in older gay men and lesbian woman than the general older adult population (30 percent versus 1 percent) and in older transgender persons compared to gender normative persons (percentages not reported). (Institute of Medicine (US) Committee on Lesbian, Gay, Bisexual, and Transgender Health Issues and Research Gaps and Opportunities, 2011) Meanwhile, the Federal Interagency Working Group (IWG) on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys has recently issued a set of working papers that identifies and evaluates current SOGI measures and provides a methodological research agenda for improving SOGI measurement. (Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys, 2016) (Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys, 2016) (Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender identity in Federal Surveys)

In early 2017 we fielded a health and tobacco survey for the population 18+ using AmeriSpeak, that included a very large oversample of the LGBT population relative to the non-LGBT population; and we used RDS network and RDS referral methods to grow the LGBT sample. We present in section 2 below a description of the RDS Network Pilot Study target population and the pilot methodology. In section 3 we discuss the experimental frameworks that apply to Network sampling and to RDS sampling. And in section 4, we present selected results from the pilot.

## **2. Description of the RDS Network Pilot Study**

Table 1 presents a comparison of incidences of “sexual identity” in two major household national probability surveys, the 2013 National Health Interview Survey and the 2008-2016 General Social Survey, and in NORC’s 2016 AmeriSpeak Panel. The incidence rates are comparable across surveys (with variability in bisexuals). Often, we are interested in assessing study outcomes for sub-groups within the LGBT population, e.g., younger or older, racial/ethnic minorities. Though AmeriSpeak has a substantial number of LGBT panel members that can be sampled for a survey (over 500), this sample size would fall short for certain studies interested in smaller subgroups of the LGBT population such as for adults >64, or for Black or Hispanic subpopulations.

In designing our pilot methodology, our primary considerations were to develop a survey-solution that would meet the following criteria:

- Develop a sampling solution that has its basis in probability sampling, while still targeting the relatively rare population of 18+ LGBT;

- Develop a comparison data set of survey data for non-LGBT adults;
- Develop an approach that supports estimation and analyses for sample representativeness and non-response bias.

Our proposed methodology attempts to satisfy these criteria by leveraging NORC's AmeriSpeak Panel as the initial sample frame for a probability based seed sample in combination with an innovative use of network sampling to generate a referral sample. The seed sample and referral sample jointly comprise the final sample of the RDS Network Pilot. We describe the pilot methodology in section 2.1, describing the AmeriSpeak Panel seed sample in section 2.2 and the RDS referral sample in section 2.3.

## 2.1 Methods of the RDS Network Pilot Study

The pilot utilized a sample selected from the AmeriSpeak panel sample as an initial base (or "seed") sample for both 18+ LGBT and 18+ non-LGBT subgroups, and the RDS nomination and referral technique to supplement the sample with additional LGBT respondents ("referrals"). As such, in addition to the nationally representative non-LGBT comparison group from the AmeriSpeak panel, the final sample combined a nationally representative probability-based sample of individuals 18 years and older who self-report as LGBT (*AmeriSpeak seed sample*) with a non-probability sample of individuals 18+ years and older who self-report as LGBT (*RDS referral sample*).

The survey length was about 10 minutes on average and completed by panelists and referrals only on the Web. Respondents were asked to complete a survey and at the end of the survey, they were asked to nominate friends and family who were LGBT to take the survey. We asked for the email addresses of the friends and family, generated a mock up of the survey invitation for the referrer to view, and send the survey invitation and survey link to all those referred.

## 2.2 AmeriSpeak Seed Sample

The AmeriSpeak panel sample relies on the NORC National Frame, an area probability sample frame constructed by NORC providing sample coverage of 97 percent of U.S. households. (Dennis, 2017) The sample design for the AmeriSpeak panel includes a two-stage recruitment process from a sample selected from the NORC National Frame. (NORC at the University of Chicago, 2017) In the first stage of recruitment, sampled households are invited via USPS mail to join AmeriSpeak either by the Web or by phone. In the second-stage of recruitment, a stratified random subsample of the nonresponders from the initial recruitment are selected and receive a second recruitment attempt via multiple modes, including face-to-face visits by field interviewers to the nonrespondents' homes to encourage participation in the panel. This multimode recruitment approach that incorporates a carefully designed nonresponse follow-up makes the AmeriSpeak panel sample representative of the U.S. household population both with and without Internet access. AmeriSpeak sample design and recruitment methodology ensures representation among important population segments, such as rural or lower income households, cell-only households, and low-income seniors age 65 and over, African Americans, Hispanics, and persons with a high school diploma or less. As of July 2017, the AmeriSpeak Panel weighted AAPOR 3 response rate was 33.5%. (Montgomery, Dennis, & Ganesh, 2016) Once recruited, panel members are sent a series of profile surveys that gather important socio-demographic characteristics in order to target panelists for surveys, reducing the need for mass screening in many cases.

Sexual orientation is one of the profile variables collected for AmeriSpeak panelists and is readily available. Thus, for the RDS Network pilot seed sample, a stratified sample by reported LGBT/non-LGBT status was selected from the panel, taking into account expected response to the survey, such that 100 LGBT and 100 non-LGBT completed interviews would be obtained. In the survey itself, panelists were asked to validate their LGBT status.

We should note that some AmeriSpeak panelists do not have access to the Internet and thus complete surveys by phone. For the pilot, considered a feasibility pilot, we excluded panelists from the sample who could not complete the survey by Web to control costs.

### 2.3 RDS Network Referral Sample

One of the strengths of RDS is that respondents invite their friends and are incentivized to encourage them to participate. Most RDS studies have been in person. Translating RDS to primarily web survey is new and relatively untested. In our implementation, the referral/invitation process is indirect in that respondents provided us with emails of friends and we sent the survey invitation to their friends.

Below, we present an overview of the RDS Network referral technique and a visual of the process is presented in Exhibit 1:

1. Select a national probability-based sample of individuals 18 and older from the AmeriSpeak panel
  - Includes a small probability-based sample of the 18+ LGBT target population
  - Includes a comparison probability-based sample of the 18+ Non-LGBT population
2. Field the Pilot Survey to the AmeriSpeak seed sample from Step 1
  - Obtain nominations of up to three other 18+ LGBT people, as well as the size of respondents' 18+ LGBT friends/family network – *Round 1 Nominations*
3. Field the Pilot survey to the referrals identified by *Round 1 Nominations*
  - Obtain nominations of up to three other 18+ LGBT people, as well as the size of respondents' 18+ LGBT friends/family network – *Round 2 Nominations*.
4. Repeat the process in Step 3 to generate up to five rounds of nominated individuals, continuing to allow up to three nominations per referee and ask about network sizes for use in calculation of weights and/or estimation.

*Exhibit 1. RDS Network Referral Protocol*



AmeriSpeak seed panelists were offered 5,000 points (1,000 points equals \$1) for completing the survey and an additional 5,000 points for each successful nomination, up to 20,000 points total. A successful nomination was a nomination that resulted in a completed LGBT interview with a nominee. Subsequent nominees were offered a \$5 Amazon gift card for each person they personally nominated, who had not already been invited to take the survey by someone else and who completed the survey. Second and later generation rounds of nominees had to provide a mailing address to receive the gift card by mail.

### 3. Experimental Referral Frameworks

Using RDS referral methods, we fielded the Pilot survey under two complimentary referral frameworks in order to investigate two estimation approaches – (A) Network estimation which is grounded in probability-based sampling and (B) RDS estimation which does not depend on probability-based sampling.

For the (A) Network estimation, the initial AmeriSpeak panel seeds and all completes that are LGBT siblings (brothers and sisters) from the **FIRST ROUND** of nominations comprise the Network sample. AmeriSpeak seed respondents were asked to report the number of siblings they had as well as the LGBT status of their siblings. Base-weights under Network estimation can be computed for the seed and referral sample using the base-weight for each unique seed respondent along with the total number of siblings referred which is asked and captured in the survey. Accordingly, the following estimates can be computed for this sample:

- a. Incidence of LGBT in population 18+
- b. Smoking and Health Status of LGBT population 18+

- c. Increase in sample size and change in design/weighting effects using the **FIRST ROUND** Network sample compared to AmeriSpeak seed sample

For (B) RDS estimation, the initial AmeriSpeak panel seeds and all completes that are LGBT friends or siblings from **ALL ROUNDS** of nominations comprise the RDS sample. RDS estimation utilizes the RDS correction for degree bias for each sample record which is  $(1/d_i)$  where  $d_i$  is self-reported number of LGBT friends/family of respondent. (Salginik & Heckathorn, 2011) (Wejnert, 2009) Accordingly, the following estimates can be computed for this sample. Note that the incidence of LGBT in the population is not appropriate since the goal of pure RDS is to generate a larger sample of only the LGBT population and not similarly grow the non-LGBT sample:

- b. Smoking and Health Status of LGBT population 18+
- c. Increase in sample size and change in design/weighting effects using ALL ROUNDS of RDS sample compared to AmeriSpeak seed sample

#### 4. Results of Pilot

We present below selected results from the RDS Network Pilot. Generally, the feasibility of the methodology was demonstrated; that we could invite a seed sample from the AmeriSpeak panel to complete a survey that included referring LGBT friends and family to take the survey and those LGBT referrals continued to refer other LGBT persons.

##### 4.1 Nominations and Referrals

AmeriSpeak seed panelists were asked how many LGBT family and friends they knew who they could ask to participate in the survey. The numbers are presented in Table 2. Almost half of the LGBT seeds knew one or more persons they could refer compared to about 20% of the non-LGBT seeds. The survey question asked of respondents from which the data were compiled is as follows:

*One of the goals of this study is to include sufficient numbers of sexual and gender minority people in the survey so that we can compare lesbian, gay, bisexual, and transgender (LGBT) persons to everyone else. In order to do this we are asking everyone who takes this survey to invite up to three LGBT persons they know to take the survey.*

*Please indicate the total number of lesbian, gay, bisexual, or transgender people who are 18 or older you know who you could send this survey to?  
ENTER NUMBER: \_*

Table 3 presents the actual number of LGBT friends referred, e.g., for which an actual email address was provided in order to send the referred person the pilot survey invitation and link. As seen in Table 3, 65 LGBT seeds that said in Table 2 they could refer actually provided 63 referrals (emails) versus 5 actual referrals from 26 non-LGBT seeds.

After three rounds of referrals, the original seed sample of LGBT completed interviews increased by approximately 20%. In all, we obtained 90 referral emails using the Web

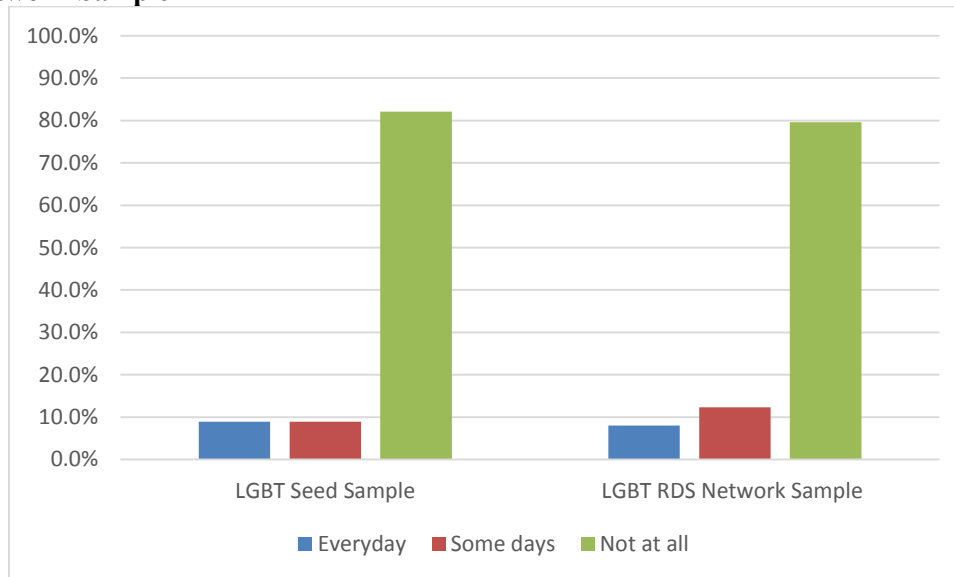
based RDS nomination and referral protocol and 28 additional completed interviews from LGBT persons that were not original AmeriSpeak panelists.

Table 4 presents the LGBT status distribution for the completed interviews from the original seed LGBT sample and the RDS Network sample comprised of the combined LGBT seed sample and referral sample. We did not find differences in the LGBT distribution between the two samples, suggesting that the referral process appears to have been fairly representative of individual statuses of Lesbian, Gay, Bisexual, and Transgender in the general population.

#### 4.2 Analysis of Smoking Outcomes

Exhibit 2 presents results from the pilot for one of the survey outcomes on vaping use. Specifically, for the question “*In the past 30 days, how often have you used e-cigarettes or other vaping products, such as vape pens, vape pipes, hookah pens, e-cigars, e-hookah, e-pipes; or eGo, Mods, APVs, RBAs; or heat-not-burn products like Revo, etc.?*”, we find the unweighted distributions between the two LGBT samples to be consistent. The addition of the RDS referral sample, did not substantively change the distribution observed in the original LGBT seed sample.

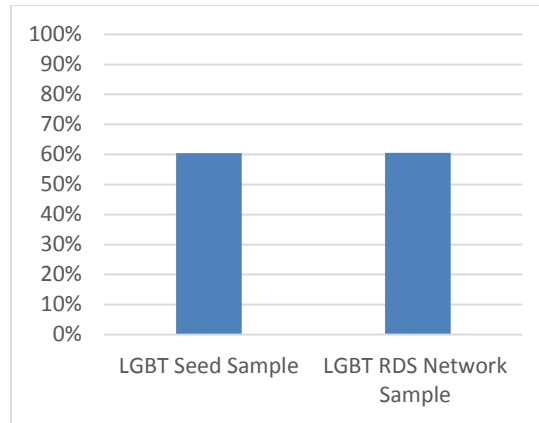
**Exhibit 2. Distribution of Vaping Use for LGBT Seed Sample and LGBT RDS Network Sample**



For the question “*Have you smoked at least 100 cigarettes in your life?*” we find the unweighted percent that reported yes in the RDS Network sample is almost exactly the same as the original unweighted AmeriSpeak seed sample – see Exhibit 3. Again, this suggests the referral LGBT sample has the desired random mechanism to it as it is so similar to the original LGBT seed sample.

**Exhibit 3. Percent Smoked at Least 100 Cigarettes in Life for LGBT Seed Sample and LGBT RDS Network Sample**





### 4.3 Estimation Methods Explored

Network sampling with a siblings counting rule was attempted through the first round of referrals in the pilot. Though the number of siblings and their LGBT status was collected successfully through the first round of referrals, the number of referred siblings was very small. The number of LGBT siblings referred from AmeriSpeak seeds was six, with five referred from the LGBT seeds and one referred from the non-LGBT seeds. Of the six referral siblings, one sibling completed the pilot survey. With such a small number of sibling completed interviews, estimation under a Network sampling framework was not practicable.

However, the RDS Network sample through all rounds of nominations under the RDS framework achieved sufficient sample size and supported application of RDS estimation methods for survey outcomes. We used utilize RDS estimation techniques included in the software package called RDSTAT<sup>2</sup>. The RDS estimator  $\hat{\mu}_f$  for a population mean of  $f$  is defined as

$$\hat{\mu}_f = \frac{1}{\sum_{i=1}^n \frac{1}{degree(X_i)}} \sum_{i=1}^n \frac{f(X_i)}{degree(X_i)}, \quad (1)$$

where  $X_1, \dots, X_n$  are the  $n$  respondents in the study and degree is the individual respondent's network number of contacts. Respondents are weighted inversely proportional to their network degree. RDS estimation does not require sample design weights for the seed sample.

Table 5 presents estimates of Vaping and Smoking utilizing RDS estimation techniques, comparing the results to unweighted estimates from the full RDS Network sample. The RDS point estimates presented in the table below are midpoints of the RDS credibility intervals computed for the Smoking or Vaping survey outcome. Somewhat larger differences are observed for Vaping behavior than for Smoking behavior.

<sup>2</sup> <http://respondentdrivensampling.org/>

## 5. Limitations

Although the pilot sample size was sufficient to prove the viability of our approach, the smaller sample sizes hindered our ability to make definitive claims about differences in survey outcomes and in comparing estimation methods. To control costs, we excluded AmeriSpeak and the referral sample who needed or preferred to take the survey by phone from our initially selected seed sample; thus, the original seed sample was representative of the population with access to the Internet. Up to three rounds of nomination were allowed in our pilot protocol; increasing the number of nomination rounds would be expected to increase the final sample size, however the percent increase per nomination may not be as robust. The incentive delivery system for gift cards and email reminder protocols in the pilot were not as optimal and efficient as we would have desired which likely impacted referral and survey completion rates.

## 6. Summary and Conclusions

We conclude that Web RDS can produce an LGBT over-sample from seeds drawn from the AmeriSpeak probability panel. We found that LGBT seeds (and referrals) were much more productive than their non-LGBT counter parts. Seeds and recruits knew many more LGBT friends and family that they said they could refer than the number of email addresses shared. We also conclude that Web network sampling with a siblings counting rule is feasible, though the number of siblings and their LGBT status collected in the pilot was quite small and is likely not a viable approach at least for the rare LGBT population.

NORC continues to develop this new approach, seeking to improve respondent facing materials and assurances, survey protocols and explanations of the importance of the research and the need for referrals. In a production model we would also hope to include the following features:

- a. Advance letters/postcards to AmeriSpeak seeds
- b. Implementing phone interviews and support and follow-up
- c. Increasing incentive levels
- d. Refining technology for prompting referrals to complete

The final RDS Network sample is a combined probability-based and non-probability based sample. As such, there are various methods for weighting and estimation of key outcomes that can be employed over and above what was investigated here. Part of next steps in this research is to continue investigating appropriate weighting and estimation methods for such samples.

## References

- Dennis, J. M. (2017, September 7). *Technical Overview of the AmeriSpeak® Panel: NORC's Probability-based Research Panel*. Retrieved from [Amerispeak.norc.org/research: http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/AmeriSpeak\\_Technical\\_Overview\\_2017\\_05\\_09.pdf?mtime=1494625611](http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/AmeriSpeak_Technical_Overview_2017_05_09.pdf?mtime=1494625611)
- Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys. (2016). *Current Measures of Sexual Orientation and Gender Identity in Federal Surveys*. Retrieved from <https://s3.amazonaws.com/sitesusa/wp-content/up>

- Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender Identity in Federal Surveys. (2016). *Evaluations of Sexual Orientation and Gender Identity Survey Measures: What Have We Learned?* Retrieved from [https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/Evaluations\\_of\\_SOGI\\_Questions\\_20160923.pdf](https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/Evaluations_of_SOGI_Questions_20160923.pdf)
- Federal Interagency Working Group on Improving Measurement of Sexual Orientation and Gender identity in Federal Surveys. (n.d.). *Toward a Research Agenda for Measuring Sexual Orientation and Gender Identity in Federal Surveys: Findings, Recommendations, and Next Steps.* Retrieved from [https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/SOGI\\_Research\\_Agenda\\_Final\\_Report\\_20161020.pdf](https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/SOGI_Research_Agenda_Final_Report_20161020.pdf)
- Gallagher, K. M., Sullivan, P. S., Lansky, A., & Onorata, I. M. (2017). *Behavioral Surveillance Among People at Risk for HIV Infection in the U.S.: the National HIV Behavioral Surveillance System.* Centers for Disease Control and Prevention.
- Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44(2), 174-199.
- Institute of Medicine (US) Committee on Lesbian, Gay, Bisexual, and Transgender Health Issues and Research Gaps and Opportunities. (2011). *The Health of Lesbian, Gay, Bisexual, and Transgender People.* Washington, DC: National Academies Press (US).
- Kates, J., & Ranji, U. (2014). *Health Care Access and Coverage for the Lesbian, Gay, Bisexual, and Transgender (LGBT) Community in the United States: Opportunities and Challenges in a New Era.* Kaiser Family Foundation. Retrieved from [health-and-access-to-care-and-coverage-for-lesbian-gay-bisexual-and-transgender-individuals-in-the-u-s](http://www.kff.org/health-equity/health-care-access-and-coverage-for-lesbian-gay-bisexual-and-transgender-individuals-in-the-u-s/)
- Montgomery, R., Dennis, J. M., & Ganesh, N. (2016). *Response Rate Calculation Methodology for Reruitment of a Two-phase Probability-based Panel: The Case of AmeriSpeak.* Retrieved from [amerispeak.norc.org/research:http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/WhitePaper\\_ResponseRateCalculation\\_AmeriSpeak\\_2016.pdf?mtime=1478878716](http://amerispeak.norc.org/research/http://d3qi0qp55mx5f5.cloudfront.net/amerispeak/i/research/WhitePaper_ResponseRateCalculation_AmeriSpeak_2016.pdf?mtime=1478878716)
- NORC at the University of Chicago. (2017, September 7). *NORC 2010 National Sample Frame.* Retrieved from NORC at the University of Chicago: <http://www.norc.org/Research/Projects/Pages/2010-national-sample-frame.aspx>
- Salginik, M. J., & Heckathorn, D. D. (2011). Sampling and Estimation in Hidden Populations Using Respondent Driven Sampling. *Sociological Methodology*, 34, 193-237.
- Schneider, J. A., Cornwell, B., Ostrow, D., Michaels, S., Schumm, P., Laumann, E. O., & Friedman, S. (2013). Network Mixing and Network Influences Most Linked to HIV Infection and Risk Behavior in the HIV Epidemic Among Black Men Who Have Sex With Men. *American Journal of Public Health*, 103(1), e28-e36.

- Schneider, J., Kozloski, M., Michaels, S., Skaathun, B., Voisin, D., & Lancki, N. (2017, January 2). Criminal justice involvement history is associated with better HIV care continuum metrics among a population-based sample of young black MSM. *AIDS*, *31*(1), 159-165.
- Sirken, M. G. (1998). A short history of network sampling. *Proceedings from the Section on Survey Research Methods*. American Statistical Association.
- Wejnert, C. (2009). an Empirical Test of Respondent Driven Sampling: Point Estimates, Variance, Degree Measures, and Out-of-Equilibrium Data. *Sociological Methodology*, *39*(1), 73-116.

**Table 1. Estimates of Incidence of Sexual Orientation**

| <b>Sexual Orientation</b>           | <b>NHIS 2013</b> | <b>GSS 2008-2016</b> | <b>AmeriSpeak 2016</b> |
|-------------------------------------|------------------|----------------------|------------------------|
| <b>Gay/lesbian</b>                  | 1.6%             | 1.7%                 | 1.7%                   |
| <b>Straight</b>                     | 96.6%            | 94.5%                | 93.3%                  |
| <b>Bisexual</b>                     | 0.7%             | 2.1%                 | 1.9%                   |
| <b>Something else</b>               | 0.2%             |                      | 0.7%                   |
| <b>Refused</b>                      | 0.6%             | 1.5%                 | 1.2%                   |
| <b>Don't Know</b>                   | 0.4%             | 0.3%                 | 1.2%                   |
| <b>Total n (unweighted)</b>         | 33,784           | 9,515                | 14,012                 |
| <b>Approximate LGB Sample Sizes</b> |                  |                      |                        |
| <b>Gay/Lesbian</b>                  | 543              | 157                  | 239                    |
| <b>Bisexual</b>                     | 221              | 196                  | 267                    |

**Table 2. Number of LGBT Persons the Seed COULD Refer (RDS Network Size)**

|  | <b>Number of LGBT Could Refer</b> | <b>LGBT Seeds</b> | <b>Percent of Total LGBT Seeds</b> | <b>Non-LGBT Seeds</b> | <b>Percent of Total non-LGBT Seeds</b> |  |
|--|-----------------------------------|-------------------|------------------------------------|-----------------------|--|--|
|  | 0                                 | 81                | 55.5%                              | 93                    | 78.2%                                  |  |
|  | 1                                 | 11                | 7.5%                               | 6                     | 5.0%                                   |  |
|  | 2                                 | 10                | 6.8%                               | 8                     | 6.7%                                   |  |
|  | 3                                 | 6                 | 4.1%                               | 6                     | 5.0%                                   |  |
|  | 4                                 | 27                | 18.5%                              | 5                     | 4.2%                                   |  |
|  | 11+                               | 11                | 7.5%                               | 1                     | 0.8%                                   |  |

|  |  |            |  |           |  |            |
|--|--|------------|--|-----------|--|------------|
| <b>Total Seeds</b>                         |  | 146        |  | 119       |  |            |
| <b>Total Potential Referrals (Up to 3)</b> |  | <b>163</b> |  | <b>58</b> |  | <b>221</b> |

**Table 3. Actual Number of LGBT Friends Referred (Up to 3 per Seed)**

| <b>Number of LGBT Referred</b>        | <b>LGBT Seeds</b> | <b>Percent of Total LGBT Seeds</b> | <b>Non-LGBT Seeds</b> | <b>Percent of Total Non-LGBT Seeds</b> |              |
|---------------------------------------|-------------------|------------------------------------|-----------------------|--|--------------|
| 0                                     | 38                | 58.5%                              | 22                    | 84.6%                                  |              |
| 1                                     | 6                 | 9.2%                               | 3                     | 11.5%                                  |              |
| 2                                     | 6                 | 9.2%                               | 1                     | 3.8%                                   |              |
| 3                                     | 15                | 23.1%                              | 0                     | 0.0%                                   |              |
| <b>Total Number of Seed Referrers</b> | 65                |                                    | 26                    |  |              |
| <b>Total Referrals</b>                | <b>63</b>         |                                    | <b>5</b>              |  | <b>68</b>    |
| <b>% Potential</b>                    | <b>38.7%</b>      |                                    | <b>8.6%</b>           |  | <b>30.8%</b> |

**Table 4. Distribution of LGBT Status for Seed Sample and RDS Network Sample**

| <b>LGBT Status</b>                 | <b>LGBT Seed Sample (n=134)</b> | <b>LGBT RDS Network Sample Thru 3 Rounds of Nominations (n=162)</b> |
|------------------------------------|---------------------------------|---|
| <b>Gay or Lesbian</b>              | 57.40%                          | 58.00%  |
| <b>Bisexual</b>                    | 37.60%                          | 37.90%  |
| <b>Something else</b>              | 1.40%                           | 1.20%   |
| <b>I don't know how to respond</b> | 3.50%                           | 3.00%   |

**Table 5. Vaping and Smoking Outcomes – Unweighted and with RDS Estimation**

| <b>Vaping/Smoking Behavior</b> | <b>LGBT RDS Sample Thru 3 Rounds (Unweighted)</b> | <b>LGBT RDS Sample Thru 3 Rounds – RDS Estimation</b> |
|--------------------------------|---|---|
| <b><i>Vaping Use</i></b>       |   |   |
| <i>Everyday</i>                | 8.0%  | 15.1%   |
| <i>Some days</i>               | 12.3%   | 32.8%   |
| <i>Not at all</i>              | 79.6%   | 59.7%   |

|   |       |       |
|---|-------|-------|
| <i>Smoked at Least<br/>100 Cigarettes in<br/>Life - Yes</i> | 60.5% | 57.6% |
|---|-------|-------|