Published in final edited form as:

Cancer Epidemiol. 2017 October ; 50(Pt B): 214–220. doi:10.1016/j.canep.2017.07.006.

The efficacy of respondent-driven sampling for the health assessment of minority populations

Grazyna Badowski^a, Lilnabeth P. Somera^{b,*}, Brayan Simsiman^c, Hye-Ryeon Lee^d, Kevin Cassel^e, Alisha Yamanaka^c, and JunHao Ren^c

^aUniversity of Guam, College of Natural & Applied Sciences, 303 University Drive, Mangilao, Guam 96923, United States

^bUniversity of Guam, Division of Communication and Fine Arts, 303 University Drive, Mangilao, Guam 96923, United States

^cUniversity of Guam, Cancer Research Center, 303 University Drive, Mangilao, Guam 96923, United States

^dUniversity of Hawaii at Manoa, Department of Communicology, 2560 Campus Road, George 329, Honolulu, HI 96822, United States

^eUniversity of Hawaii Cancer Center, 701 Ilalo St., Honolulu, HI 96813, United States

Abstract

Background—Respondent driven sampling (RDS) is a relatively new network sampling technique typically employed for hard-to-reach populations. Like snowball sampling, initial respondents or “seeds” recruit additional respondents from their network of friends. Under certain assumptions, the method promises to produce a sample independent from the biases that may have been introduced by the non-random choice of “seeds.” We conducted a survey on health communication in Guam’s general population using the RDS method, the first survey that has utilized this methodology in Guam. It was conducted in hopes of identifying a cost-efficient non-probability sampling strategy that could generate reasonable population estimates for both minority and general populations.

Methods—RDS data was collected in Guam in 2013 ($n = 511$) and population estimates were compared with 2012 BRFSS data ($n = 2031$) and the 2010 census data. The estimates were calculated using the unweighted RDS sample and the weighted sample using RDS inference methods and compared with known population characteristics.

*Corresponding author at: University of Guam, Cancer Research Center, 303 University Drive, Mangilao, Guam 96923, United States. lpsomera@triton.uog.edu (L.P. Somera).

Conflict of interest

The authors declare no potential conflicts of interest.

Contribution of authors

Study conception and design: Grazyna Badowski, Lilnabeth P. Somera, Hye-ryeon Lee, Kevin Cassel, Brayan Simsiman.

Acquisition of data: Brayan Simsiman, Alisha Yamanaka, Grazyna Badowski, Lilnabeth P. Somera.

Data Analysis: Grazyna Badowski, Brayan Simsiman, Alisha Yamanaka, JunHao Ren.

Interpretation of data and writing the article: Grazyna Badowski, Lilnabeth P. Somera, Brayan Simsiman, Hye-ryeon Lee, Kevin Cassel, Alisha Yamanaka.

All authors reviewed and approved the final manuscript.

Results—The sample size was reached in 23 days, providing evidence that the RDS method is a viable, cost-effective data collection method, which can provide reasonable population estimates. However, the results also suggest that the RDS inference methods used to reduce bias, based on self-reported estimates of network sizes, may not always work. Caution is needed when interpreting RDS study findings.

Conclusions—For a more diverse sample, data collection should not be conducted in just one location. Fewer questions about network estimates should be asked, and more careful consideration should be given to the kind of incentives offered to participants.

Keywords

Respondent-driven sampling; Health communication; Guam; Chamorro; Federated States of Micronesia; Micronesians; Pacific Islanders

1. Introduction

Despite overall health improvement, significant health disparities persist in minority populations [1]. Efforts to reduce these disparities are hindered by a lack of good surveillance data. Obtaining probability samples representing various minority populations through the preferred method of random digit dialing (RDD) is not only cost-prohibitive, but also ineffective in reaching many minority groups. Widely used for national surveillance studies such as the Behavioral Risk Factor Surveillance Survey (BRFSS), RDD is impractical for populations of small size (hidden populations), which do not have a proper sampling frame [3].

The lack of scientific surveillance data has serious consequences. It makes designing effective programs for underserved populations nearly impossible and limits government or private funding opportunities for health programs and interventions. Thus, a cost-effective means of obtaining reliable data for minority populations is essential.

In this study we propose respondent-driven sampling (RDS) as an efficient, cost-saving alternative to RDD. Using data collected with the RDS method in Guam, we compare the resulting data with the BRFSS data collected by RDD.

1.1. Background

Guam's multiethnic population offers a unique opportunity to learn about the health information needs and practices of Americans of Pacific Island ancestry, a largely underserved and underrepresented segment of the national population. A United States territory in the western Pacific with an area of 212 square miles and a population over 170,000 as of 2016, Guam is composed of numerous cultural groups and languages [4] including Chamorros, Filipinos, and other Pacific Islanders. Good data on Americans of Pacific Islander ancestry (API), which critical for developing effective health campaigns, establishing baseline information, measuring the impact of these campaigns, and comparing local with national and international patterns in health behaviors, is very limited. This makes Guam's population ideal for testing the respondent-driven sampling (RDS) methodology, since RDS targets hard-to-reach or hidden populations.

Data which provide a more precise and culturally accurate picture of the community is important for developing more effective cancer prevention and control programs that are tailored to the target population. While API comprise only 4% of the U.S. population, it is one of the fastest-growing groups in America (2010 U.S. Census). Yet, there is limited data on API, among whom the incidence of certain cancers and certain noncommunicable diseases (NCDs) are higher than the national average [5,6]. Moreover, the labels for the distinct groups which fall under the API umbrella term continue to evolve in national data. The terms “Guamanian” and “Samoan” were only included with “Hawaiian” on U.S. census forms since 1980.

Data for Guam is further complicated by the presence of citizens from the Federated States of Micronesia (FSM), who can live and work in U.S. territory by virtue of the Compact of Free Association (COFA) with the United States. In the 2010 Census, 10% of the Guam population were FSM residents, while another 9.4% reported that they were “mixed.” Given these population complexities in Guam and its diverse cultural and linguistic backgrounds, acquiring a good probability sample of Guam residents poses unique challenges in terms of cost [2]. Consequently, communication interventions designed to address health disparities may be based on no data or data which do not capture an accurate picture of the target population.

Since it is substantially cheaper than probability sampling methods such as RDD, if RDS can successfully generate a sample that is statistically representative of the various ethnic groups in Guam, it will be a significant step in satisfying the needs for good health data. We intend to investigate if RDS method can produce stable estimates for key health outcomes comparable to those from BRFSS data collected through the more expensive RDD method.

1.2. Respondent-driven sampling

Respondent-driven sampling (RDS) is a fairly new form of chain referral sampling developed by Douglas Heckathorn [3,7], which uses chain-referrals that progress through a series of recruitment waves until equilibrium (when composition of the ultimate sample is independent of the initial sample) is reached. This advanced method of sampling combines “snowball sampling” with the employment of mathematical modeling, stochastic Markov chain modeling, that weights the sample to compensate for the non-random sample collection.

RDS differs from snowball sampling, in several important ways. First, while snowball sampling only gives incentives for participation, RDS has a dual incentive system: for participating as well as for recruiting others into the study. Second, RDS asks subjects to actually recruit their peers into the study, compared to simply identifying them. This helps in two ways: individuals who might be reluctant to give a researcher the name of a peer might, nonetheless, recruit that peer. In addition, people, who might refuse to participate when approached by a researcher, will agree to a peer’s invitation. This system creates larger personal networks to recruit from, instead of relying on subjects with smaller networks. However, the limit placed on the number of recruits, typically three per individual, ensures that recruitment is not biased by reliance on a few individuals who are more effective recruiters. Most importantly, RDS can produce unbiased population estimates, unlike

snowball sampling. It does so by taking into the account that study participants were not recruited randomly and uses statistical weights based on the participants' network size (i.e., the number of people that the participants know who would be eligible for the study) and recruitment patterns (who recruited whom). As recruitment continues across waves, equilibrium that results in a sample independent of the characteristics of the initial seeds can be attained.

In this study, the RDS recruiting method was used among members of Guam's general population to assess if this method can be used to obtain a statistically representative population based sample comparable to that obtained through RDD.

2. Materials and methods

Data used for this study were collected during a period of twenty-three (23 days) from February 1 to March 6, 2013 by the Guam Cancer Research Center. All adults (18+ years old) from the general Guam population were eligible to participate in the paper-and-pencil survey that included questions about their health information seeking and health behaviors. The recruitment center was located on the campus of the University of Guam, in the centrally located village of Mangilao.

A total of fourteen "seeds" varying in gender, age, ethnicity, background and geographical representation were selected by convenience from the target population. Table 1 summarizes their characteristics. The ethnic distribution of these seeds includes 43% Chamorro (6/14), 35% Filipino (5/14), 7% Chuukese (1/14) and 15% other ethnic groups which closely reflects the current distribution of Guam's multiethnic population that according the 2010 U. S. Census includes 37.3% Chamorro, 26.3% Filipino, 7% Chuukese and 30% other ethnic groups (CIA World Fact Book <https://www.cia.gov/library/publications/the-world-factbook/geos/gq.html>).

Each of the 14 seeds completed a paper-and-pencil survey. They were provided with a coupon they can use to recruit up to three (3) their acquaintances to participate in the survey, as illustrated in Fig. 1. The coupon included key information such as location, expiration date, and contact information.

Participants were informed that their potential recruits will receive the incentive if they came in and completed the survey within fourteen (14) days. Seeds and recruits (Wave 1) were offered dual incentive for participation and recruitment. All participants were given a \$25 gas coupon for survey completion and an additional \$20 gas coupon for each person recruited. The participants were allowed to recruit a maximum of three recruits (Wave2) to avoid the issue of "professional recruiting" [3]. This process continued for all succeeding waves until the target sample was reached.

RDS estimation requires information on the network size. We used the network-size questions used in other RDS studies [8,9] which included the following network questions:

N1: How many people do you know on Guam personally?

N2: With how many people from Guam did you interact via phone, email, meeting in person in the past 30 days?

N3: With how many people from Guam do you discuss important matters with?

N4: What is the number of people on Guam that you think you could potentially recruit for the study?

N5: How many CLOSE FRIENDS do you have on Guam? N6: How many FRIENDS do you have on Guam?

N7: How many ACQUAINTANCES do you have on Guam? Coupon distribution and tracking were achieved using RDS

Coupon Manager 1.0. Data processing was performed by IBM SPSS Version 21. Once data collection was complete, information from the coupon manager and survey were merged and analyzed using RDSAT v7.1.38 [10]. RDS-1 point estimators and 95% confidence intervals were calculated using 15,000 bootstraps iteration [11–14]. RDS-2 point estimators were calculated using RDS Analyst [15]. Recruitment trees of each seed were generated utilizing GraphViz [16].

Simple respondent-driven sampling sample proportions and respondent-driven sampling estimates were calculated for each of the network size questions separately, with and without the outliers. Root mean squared errors were calculated for the differences between the population proportions and the RDS-1 and RDS-2 estimates, for each variable and in total and for each network size estimate to determine which network question performed the best.

3. Results

A total of 511 people, including the 14 seeds, were recruited and completed the survey. The recruitment distribution by days is shown in Fig. 2A. The cumulative number of recruits over time is shown in Fig. 2B. After the fifteenth (15th) day, no new coupons were distributed, as it was expected that the sample size would be met from previously scheduled appointments. Out of the fourteen (14) original seeds, only eleven (11) seeds recruited people into the study. Of all the recruits, 92.4% ($n = 459$) came from three (3) seeds, with more than half of the recruits coming from a single seed (52.1% of the overall sample, $n = 497$, excluding seeds). Another seed recruited 132 persons or 26.6% of the sample, while the third brought in 68 recruits or 13.7% of the sample. Four (4) other seeds recruited three (3) people each, 3 seeds recruited 2 people, and the remaining four (4) seeds each recruited one (1) person. The total number of recruits originating from the 11 seeds ranged from 1 to 259.

The number of waves from the eleven (11) recruiting seeds ranged from 1 to 13 waves. The highest recruitment occurred in the fifth (5th) wave, which produced 14.1% of the total recruitment, excluding the seeds. Waves 4–8 accounted for 64.8% of the total recruitment (excluding seeds) and averaged about 64.4 recruits per wave. The results are presented in Fig. 2C.

To evaluate the efficacy of respondent driven sampling, we compared estimates from the RDS survey on health communication with total population data from the 2010 Guam

Census and the 2012 Behavioral Risk Surveillance Survey (BRFSS) that uses conventional RDD sampling methods. Tables 2a and b present the comparisons between the 2010 Census data, 2012 BRFSS, RDS sample proportion and RDS-1 and RDS-2 estimates. The sample proportions were close to the population's proportions for most characteristics. The RDS sample was largely representative of the total population by sex, ethnicity, socioeconomic status and geographic location. However, the sample overrepresented young adults age 18–34 and with some post high school education.

Comparison of BRFSS estimates with unweighted and weighted RDS estimates was also conducted. The RDS estimates of some risk behaviors (smoking, BMI) were close to BRFSS estimates, when the unweighted sample was used. The distributions of BMI and smoking status obtained in our study were not statistically different from BRFSS 2012 distributions (chi-square tests, p-value = 10.8 and 12.3 respectively)

Respondent-driven sampling statistical inference methods did not reduce the age and education biases regardless of the network size question used for the analysis. Taking out the outliers from the network size did not improve the estimates neither. This is consistent with findings from other studies [17]. The root mean squared error for the difference between the population proportions and the respondent driven sampling estimates ranged from 11.4% when the network question N3 was used in calculations to 14.5% for the network question N7.

One possible explanation for this finding is that the population may not have accurately reported their network size. The distributions of the reported network size were right skewed with most values being multiples of five (5) and ten (10). The proportion of values that were multiples of five ranged from 69% to 92.1%. The proportion of values that were multiples of ten ranged from 45.4% to 85.1%. This suggests that some of the responses may not have been very thoughtful. N3 and N5 had the lowest percentage of multiples of five (5) and ten (10) (69% and 45.5%; 66.5% and 50.3%, respectively). Some of the respondents gave the same answer for all network questions.

The analysis of the dynamics of the recruitment networks suggests an explanation for the under-recruitment of older people. From the recruitment networks in Fig. 3, we can see that older people (50 or over) were at the end of the recruitment chain. Fig. 2C shows that most of them completed the survey after the eighth (8th) wave. They were also more likely to come in later in the day and on weekends. Furthermore, the numbers of days between recruiters' survey and their recruits' survey is significantly different for different age groups (chi-square test, p-value = 0.02). There were more days from the time older recruiters completed their surveys to the day when their recruits completed theirs. These differences are shown in Fig. 2D.

4. Discussion

This study is the first known use of the respondent-driven sampling methodology in health communication research in Guam. It produced a largely representative sample of Guam

population for many variables, including ethnicity, income, and some selected behavioral risk factors.

The use of RDS in this study demonstrates that it can potentially be a viable method for recruiting research participants, as far as obtaining quick estimates is concerned. However, the fact that it only produced comparable estimates for some health outcomes suggests a cautionary note regarding its attendant biases.

The target sample ($N = 500$) was surpassed in twenty-three (23) days with the final sample of five hundred eleven (511). In fact, the extremely positive response created some initial challenges as the researchers tried to manage the respondent participation flow. More controlled participation resulted as eligible recruits who received coupons had to schedule appointments. Schedules for surveys were based on the limited hours of the research staff. Thus, the number of surveys completed on each day was constricted.

One major exception to the representativeness of the sample was the overrepresentation of young adults (ages 18–34) and participants with some post high school education. This bias in our sample might have affected estimates of variables that are known to be influenced by age and education level. For example, with a smaller number of young adults with post high school education in the sample, our data may have had higher level of smoking prevalence and higher level of overweight, thus producing results closer to BRFSS data. The most plausible explanation for the sample bias may have been the recruitment center's location, which was on the university campus. Therefore, it was easier for students to recruit each other and participate in the study. On the other hand, older people took longer to come to the center. Due to the budget constraints, the sample size was limited to 512. However, the analysis of the recruitment networks suggests that if we had continued with sampling process, we might have obtained a better representation of the older age group.

The dual incentive appeared to be a factor in achieving the study sample size in such a short time period. Possible selection bias may have occurred since respondents could potentially receive \$85 in gas cards if they participated in the survey and recruitment. It was pointed out [17] that the incentive offered for participation “may look small to some people and big to others.” Evidently, the RDS incentive was more attractive to younger people, who responded more quickly.

Moreover, gas prices going up the week before the study started may also have made the gas cards offered as incentives more attractive than they would have been otherwise. It may also explain why the younger population was oversampled, and resulted in a bias by age and socioeconomic status, as college students with less disposable incomes found the gas card incentives more appealing than older adults.

The survey questions regarding network characteristics resulted in inaccurate estimates since the participants over-reported values that were in multiples of 5 or 10. Since the methodology relies heavily on self-reported estimates of their network sizes, the survey questions and the kind of responses they generated did not met the assumptions behind the RDS method, and suggests that RDS sampling methods may have increased certain biases.

In future studies, fewer questions about network estimates should be asked, as more questions did not seem to lead to more precise answers.

Our study produced some promising results regarding the use of the RDS as a data collection method, given its relative ease and lower cost of administration. At the same time, the results also identify several factors that could limit the effectiveness of the RDS statistical-inference methods in yielding more precise network estimates. Future studies should therefore take into account the kind of incentives offered for participation in an RDS study, and consider how valuable or attractive they may be to the target population, as well as the location of the research venue and its proximity to the target population. As we found, these factors could affect the diversity and representativeness of the sample. In this case, increasing the number of venues may have resulted in a more diverse sample. In conclusion, while the relatively lower cost of RDS makes it an attractive alternative to other methodologies, our results indicate that it comes with its own attendant biases. More research using the RDS method is needed before definitive conclusions can be made regarding its efficacy in producing population estimates that are comparable to samples recruited using the RDD method.

Acknowledgments

This research was supported by a U54 Minority Institution/Cancer Center Partnership Grant from the National Cancer Institute (Nos. 5U54CA143727 and 5U54CA143728).

We would like to acknowledge Megan Inada, Vejohm Torres, Dani Reyes, and Joshua Tobias for their assistance in the research project.

1.4 References

1. Angeles, J., Somers, SA. From Policy to Action: Addressing Racial and Ethnic Disparities at the Ground-Level. Center for Health Care Strategies, Inc; Hamilton, NJ: 2007.
2. Heckathorn D. Respondent-driven sampling: a new approach to the study of hidden populations.pdf. Soc Probl. 1997; 44:174–199. <http://www.respondentdrivensampling.org>.
3. Palafox NA, Hixon AL. Health consequences of disparity: the US affiliated pacific islands. Aust Psychiatry. 2001; 19:S84–S89. doi:<http://dx.doi.org/10.3109/10398562.2011.583072>.
4. Somera L, Lee H-R, Badowski G, Cassel K. Health information seeking, source trust, and culture: a comparative analysis of health information trends and needs between Guam and the United States. J Health Commun. 2016; 21:469–478. doi:<http://dx.doi.org/10.1080/10810730.2015.1095822>. [PubMed: 26983674]
5. The Centers for Disease Control and Prevention, CDC. 2013
6. Haddock RL, Whippy HJ, Talon RJ, Montano MV. Ethnic disparities in cancer incidence among residents of Guam. Asian Pac J Cancer Prev. 2009; 10:57–62.
7. Heckathorn, D., Semaan, S., Broadhead, R., Hughes, J. Extensions of Respondent-Driven Sampling. 2002. p. 18-25.doi:<http://dx.doi.org/10.1023/A:1014528612685>
8. McCarty C, Killworth PD, Bernard HR, et al. Comparing two methods for estimating network size. Hum Organ. 2001; 60:28–39.
9. McCormick TH, Salganik MJ, Zheng T. How many people do you know?: Efficiently estimating personal network size. J Am Stat Assoc. 2010; 105:59–70. doi:<http://dx.doi.org/10.1198/jasa.2009.ap08518>. [PubMed: 23729943]
10. Volz, E., Wejnert, C., Cameron, C., Spiller, M., Barash, V., Degani, I., Heckathorn, DD. Respondent-Driven Sampling Analysis Tool (RDSAT) Version 7.1. Cornell University; Ithaca, NY: 2012.

11. R. Incorporated. RDS Analysis Tool Manual. 2006. <http://www.respondentdrivensampling.org>
12. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Probl. 1997; 44:174–199.
13. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. Soc Meth. 2004; 34:193–240.
14. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. Soc Probl. 2002; 49:11–34.
15. Handcock, M., Fellows, I., Gile, K. RDS Analyst: Software for the Analysis of Respondent-Driven Sampling Data, Version 0.42. 2014. <http://www.hpmrg.org>
16. Gansner ER, North SC. An open graph visualization system and its applications to software engineering. Softw Pract Exp. 1999; 30:1203–1233. doi:[http://dx.doi.org/10.1002/1097-024X\(200009\)30](http://dx.doi.org/10.1002/1097-024X(200009)30).
17. McCreesh N, Frost SDW, Seeley J, et al. Evaluation of respondent-driven sampling. Epidemiology. 2012; 23:138–147. doi:<http://dx.doi.org/10.1097/EDE.0b013e31823ac17c>. [PubMed: 22157309]

Fig. 1.
Respondent-driven sampling recruitment process.

Fig. 2.

Summary of respondent-driven sampling recruitment in Guam. A: The number of recruits by day; B: The cumulative recruitment over time (including seeds); B: The total number of recruits by each seed (excluding seeds); C: The number of recruits per wave by each seed (including seeds); D: Number of days between recruiter's and recruit's participation by age groups.

Fig. 3.
Three largest recruitment networks showing gender and age groups. Seeds are shown at the top of each recruitment network.

Seeds' Characteristics.

Table 1

Seed Number	Age	Gender	Ethnicity	Village	Reported Network size (N3)	Number of recruits	Number of waves	Total number of recruits
1	36	Female	Palauan	Mangilao	5	3	8	132
2	58	Male	Pohnpeian	Talofofo	2	1	4	11
3	39	Female	Chamorro	Talofofo	500	1	3	11
4	21	Female	Chuukese	Mangilao	20	3	2	7
5	24	Male	Filipino	Barrigada	30	3	13	259
6	47	Female	Filipino	Dededo	20	1	1	1
7	23	Female	Chamorro	Yona	10	0	0	0
8	23	Female	Filipino	Yigo	8	0	0	0
9	57	Male	Filipino	Yigo	15	3	8	68
10	61	Male	Filipino	Merizo	10	0	0	0
11	63	Male	Chamorro	Dededo	5	1	2	2
12	28	Male	Chamorro	Yona	30	1	1	1
13	37	Female	Chamorro	Yona	10	1	1	3
14	55	Female	Chamorro	Agana Heights	5	1	1	2

Table 2a
Population Proportion, Sample Unweighted Proportions, and RDS-1 and RDS-2 Estimates.

Estimate (95% CI)				
Unweighted sample				
	proportion	RDS-1	RDS-2	2010 Census Data
Gender				
Male	43.6	43.1 (36.8–49.4)	43.0 (36.11–49.89)	51.0 –
Female	56.4	56.9 (50.6–63.2)	57.0 (50.11–63.9)	49.0 –
Age groups (years)				
18–34	66.9	67.2 (58.7–75.1)	73.5 (58.5–73.49)	36.3 –
35–54	22.4	22.8 (16.6–29.3)	29.9 (18.05–29.85)	31.9 –
55+	10.8	10.0 (5.9–15.0)	14.4 (5.72–14.38)	31.8 –
Ethnicity				
Chamorro	51.6	49.3 (40.4–57.7)	44.3 (35.69–52.29)	38.46 –
Filipino	30.8	28.6 (21.1–36.8)	33.2 (25.09–41.23)	31.5 –
White	2.9	3.2 (1.3–5.1)	3.1 (0–7.7)	8.86 –
Asian	4.1	5.4 (2.3–9.3)	6.0 (2.35–9.58)	7.64 –
Pacific Islander	10.0	12.8 (6.9–20.0)	12.8 (9.78–15.87)	7.13 –

Table 2b
Comparison of BRFSS estimates with Sample Unweighted Proportions, RDS-1 and RDS-2 Estimates.

Estimate (95% CI)				
	Unweighted proportion	RDS-1	RDS-2	2012 BRFSS
BMI				
underweight	2.7	2.9 (1.2–5.2)	3.1 (0.07–5.25)	2.9 (1.9–4.0)
normal weight	36.2	37.2 (31.0–43.2)	36.8 (30.50–43.16)	35.6 (32.8–38.4)
overweight	28.9	29.0 (23.2–35.2)	29.1 (23.46–34.77)	32.4 (29.6–35.2)
obese	31.7	30.9 (25.3–37.1)	31.0 (25.13–36.93)	29.1 (26.3–31.9)
Smoking Status				
current smoker	19.4	19.6 (14.3–24.7)	20.5 (14.69–25.8)	25.8 (23.1–28.5)
former smoker	15.9	15.5 (10.8–21.9)	16.1 (11.19–20.95)	13.5 (11.6–15.4)
never smoked	64.7	64.9 (57.5–71.6)	63.7 (56.37–71.00)	60.7 (57.8–63.6)
Education				
Less than HS	9.3	10.6 (6.8–15.6)	7.0 (3.54–10.53)	20.5 (17.6–23.3)
HS or GED	28.5	30.3 (23.7–35.6)	33.0 (26.41–39.67)	33.9 (31.3–36.5)
Some Post HS	41.3	41.4 (34.6–49.4)	39.9 (32.70–47.00)	19.6 (17.4–21.8)
College	20.9	17.8 (12.2–23.6)	20.1 (15.97–24.18)	26.0 (23.6–28.4)
Income				
<\$15,000	23.9	29.5 (21.8–36.6)	25.9 (18.78–32.94)	20.3 (17.7–23.2)
\$15,000–\$34,999	25.0	26.3 (20.8–32.2)	26.5 (20.2–32.80)	35.9 (32.9–39.1)
\$35,000–\$49,999	15.9	15.3 (10.5–20.7)	15.6 (9.64–21.55)	14.3 (12.3–16.6)
\$50,000–\$74,999	17.1	14 (10.1–18.6)	14.7 (9.37–20.64)	13.1 (11.2–15.2)
\$75,000+	18.1	15 (10.2–20.7)	17.3 (13.38–21.29)	16.4 (14.3–18.7)