Finding Needles in a Haystack: A Methodology for Identifying and Sampling Community-Based Youth Smoking Cessation Programs
Sherry Emery, Jungwha Lee, Susan J. Curry, Tim Johnson, Amy K. Sporer, Robin Mermelstein, Brian Flay and Richard Warnecke

*Eval Rev* 2010 34: 35
DOI: 10.1177/0193841X09355258

The online version of this article can be found at:
http://erx.sagepub.com/content/34/1/35
Finding Needles in a Haystack: A Methodology for Identifying and Sampling Community-Based Youth Smoking Cessation Programs

Sherry Emery, MBA, PhD,1 Jungwha Lee, PhD, MPH,2 Susan J. Curry, PhD,3 Tim Johnson, PhD,4 Amy K. Sporer, MS,1 Robin Mermelstein, PhD,1 Brian Flay, DPhil,5 and Richard Warnecke, PhD1

Abstract
Background: Surveys of community-based programs are difficult to conduct when there is virtually no information about the number or locations of the programs of interest. This article describes the methodology used by

1 Institute for Health Research and Policy, University of Illinois at Chicago
2 Department of Preventive Medicine, Northwestern University, Chicago, Illinois
3 College of Public Health, University of Iowa
4 Survey Research Laboratory, University of Illinois at Chicago
5 College of Health and Human Sciences, Oregon State University

Corresponding Author:
Sherry Emery, Institute for Health Research and Policy, University of Illinois at Chicago, 1747 W. Roosevelt Rd., Suite 558, Chicago, IL 60608.
Email: http://ihrp-web.ihrp.uic.edu/
the Helping Young Smokers Quit (HYSQ) initiative to identify and profile community-based youth smoking cessation programs in the absence of a defined sample frame. **Methods:** We developed a two-stage sampling design, with counties as the first-stage probability sampling units. The second stage used snowball sampling to saturation, to identify individuals who administered youth smoking cessation programs across three economic sectors in each county. Multivariate analyses modeled the relationship between program screening, eligibility, and response rates and economic sector and stratification criteria. Cumulative logit models analyzed the relationship between the number of contacts in a county and the number of programs screened, eligible, or profiled in a county. **Results:** The snowball process yielded 9,983 unique and traceable contacts. Urban and high-income counties yielded significantly more screened program administrators; urban counties produced significantly more eligible programs, but there was no significant association between the county characteristics and program response rate. There is a positive relationship between the number of informants initially located and the number of programs screened, eligible, and profiled in a county. **Discussion:** Our strategy to identify youth tobacco cessation programs could be used to create a sample frame for other nonprofit organizations that are difficult to identify due to a lack of existing directories, lists, or other traditional sample frames.

**Keywords**
community health promotion programs, program evaluation, youth smoking cessation, organizational research

**Introduction**

How does one undertake research about community-based programs, about which there is so little information that we do not know how many there are, or where to look? The literature on organizational or institutional surveys might provide some guidance, but most work on organizational survey and sampling methods has emphasized surveys of well-defined and easily locatable entities, such as governmental, business, and educational organizations. There is also a large literature describing common approaches available for finding difficult-to-locate populations or individuals, including probability-based methods such as population screening and network (or multiplicity) sampling (Birnbaum and Sirken 1965; Kalton and Anderson 1986; S. Thompson and Serber 1996) and nonprobability
methods such as snowball sampling (Sudman and Kalton 1986) and respondent-driven sampling (Heckathorn 1997 and 2002; W. Thomson et al. 2002). The strengths and limitations of each approach are well recognized, and each is commonly used in practice. There has, however, been very little published research on how to conduct research on rare or difficult-to-find community-based programs. Such research is particularly difficult to conduct when there is virtually no information available about the number or locations of the organizations or programs of interest.

The Helping Young Smokers Quit (HYSQ) initiative faced just such a challenge when it set out to evaluate and inform best practices in community-based youth smoking cessation programs. Before undertaking a national evaluation of youth tobacco cessation programs, it was necessary to determine whether there were a sufficient number of replicable programs to evaluate. Thus, we sought to establish the prevalence and location of community-based youth smoking cessation programs in the United States and describe the types of services they were providing.

There were a number of inherent challenges in identifying cessation programs for youth. Because such programs are not subject to licensing or other regulations, they could not be identified through registries, tax records, or other means often used in organizational sampling. Public health programs such as youth smoking cessation interventions may exist within a variety of larger organizational structures—such as schools, health departments, hospitals, or other youth organizations—that can camouflage their presence. Such programs likely receive funding via a variety of mechanisms and a variety of public and private sources. As a result, it is also difficult to generate reasonable projections regarding the numbers of such organizations that one might expect to identify in any given geographic area.

This article describes HYSQ’s methodology for identifying and profiling community-based youth smoking cessation programs in the absence of a defined sample frame. While our focus was on finding a specific type of community-based health organization, our experience and methodology is directly translatable to research involving other small, not-for-profit organizations and public health organizations. We begin by briefly describing a key informant and snowball sampling strategy developed to systematically search for and identify youth smoking cessation programs.1 We next consider which elements of this strategy were most productive and how each contributed to the composition of our sample. We conclude with a consideration of the strengths, limitations, and lessons learned from this approach, along with a set of recommendations for conducting similar surveys of small public health programs.
Method

Our strategy for seeking and identifying community-based youth smoking cessation programs in the United States began with a clear operational definition of the target sample. We then developed a two-stage sampling design, with counties as the first-stage probability sampling units. The second stage used snowball sampling to identify individuals who administered youth smoking cessation programs. Cessation programs were profiled as they were identified. The goal of the snowball sampling strategy was to identify all contacts at the local level, who were knowledgeable about youth smoking cessation programs in their communities, or who could lead to such persons. Snowball sampling progressed through two “tiers” and ended with the identification of a program informant who administered a youth smoking cessation program in the community. The intent of this two-stage process was to contact the entire population of individuals within each county, who would have knowledge of, or be an administrator of, a youth smoking cessation program.

The critical elements of our methodology, described in detail below, were to:

- identify a proxy sample frame from which to search;
- stratify the sample frame by key variables of interest;
- use prior research and practical knowledge of youth smoking cessation treatment to identify likely sectors within the units of the sample frame from which to begin the search for programs;
- develop lists of Tier 1 key informants within each sector, using publicly available information;
- snowball sample for additional informants and program administrators, referred to as Tier 2 contacts; and
- interview potential program administrators to screen for program eligibility for inclusion in the survey.

Operational Definition of a Youth Smoking Cessation Program

Prior to beginning the search for youth smoking cessation programs, we operationally defined these programs based on five criteria. Eligible programs (a) were established at least 6 months prior to being contacted by the HYSQ evaluation; (b) were provided direct tobacco cessation services; (c) had at least half of their participants aged 12–24 years; (d) were
not currently part of a research initiative; and (e) did not focus on pregnant smokers.

**Stage 1: Sampling Counties**

*A proxy sample frame.* Absent a comprehensive list of youth smoking tobacco cessation programs across the United States, we needed to identify a working, or proxy sampling, frame that would enable us to organize our search around clearly defined units of analysis. Because our research aimed to determine the prevalence and location of community-based youth smoking cessation programs in the United States, we focused on how to operationally define “community.” Counties are a natural unit of political and economic activity, as well as geography, and thus served our purpose well. The overall proxy sampling frame comprised 2,453 counties (3,124 counties in the United States minus 689 with populations less than 10,000 deemed unlikely to yield programs).

*Stratification.* We stratified counties by four characteristics, which we hypothesized would be associated with program availability: urbanization (metropolitan statistical area [MSA] vs. non-MSA); socioeconomic status (SES, high vs. low); young adult smoking prevalence (high vs. low); and per capita tobacco control expenditures (high, medium, low). We expected to find more programs in counties that were urban, with higher SES and with higher tobacco control expenditures. We used young adult smoking prevalence as a proxy for youth smoking prevalence and hypothesized that higher smoking prevalence would be associated with lower program availability. Urbanization and SES were defined at the county level; smoking prevalence and per capita tobacco control expenditures were defined at the state level because county data were unavailable.

Urban counties lay within a MSA as defined by the U.S. Census Bureau. Low SES counties were those in which more than 20% lived below the federally defined poverty level, based on 2000 Census data. High smoking prevalence counties included all counties in a state where the current smoking prevalence for 18- to 24-year olds was above the national median of 31%, based on the 2000 Behavioral Risk Factor Surveillance System. Three levels of weighted per capita tobacco control expenditures were defined based on state-level data obtained by RTI International (Farrelly, Pechacek, and Chaloupka 2003). States were rank-ordered by their weighted average expenditures and grouped by tertiles.
The four stratification criteria produced 24 strata. Some strata contained very few counties and were subsequently collapsed, resulting in a final set of 17 strata. Based on cost limitations and expert judgment about the likely availability of programs, we set a target of 400 interviews with eligible youth smoking cessation programs and used that target to guide the number of counties to include in the second stage of sampling. Aiming for a total of 408 counties, we sampled unequally across strata, presuming a greater likelihood of finding programs in states with higher per capita expenditures for tobacco control. Eighteen counties were selected from each of the six “low-expenditure” strata, 24 counties were selected from each of the five “medium-expenditure” strata, and 30 counties were selected from each of the six “high-expenditure” strata.

Within each stratum, counties were randomly selected, without replacement, in proportion to the size of the youth population (aged 10–24 years) in the county, anticipating that we were more likely to find youth targeted programs in counties with a higher youth population. Forty-nine states are represented in the 408 counties.

**Identifying Sectors**

Within each county, we identified three organizational sectors in which we expected to find youth smoking cessation programs, based on prior research and experience with youth and adult smoking cessation treatment programs:

1. county departments of health;
2. county-based school districts; and
3. major voluntary organizations comprising the American Cancer Society, American Lung Association, and the American Heart Association.

We also accepted a list of trained providers from an organization that disseminated a youth smoking cessation program as a fourth source of key informants that we label as an “other” sector. The identification of sectors further focused our search for key informants.

**Stage 2: Snowball Sampling**

**Tier 1 key informants.** Replicable protocols directed the identification of first-tier informants from each of the sectors. These Tier 1 informants were typically the lead administrators of each sector. For example, we searched...
for the director of county health departments; the director of school health or an equivalent position within school districts; and the president or executive director of the major voluntary organizations. If no contacts could be identified through county-level domains, state-level organizations were contacted. Tier 1 key informants were identified from Web sites, telephone directories, and (in the case of the Other sector) lists of individuals likely to be knowledgeable about youth cessation treatment.

**Tier 2 contacts.** Tier 2 contacts were identified by asking Tier 1 key informants for contact information of anyone whom they believed to be knowledgeable about youth tobacco cessation in their county. Snowball sampling continued until no new contacts were identified in a given county.

**Identifying youth smoking cessation programs.** All Tiers 1 and 2 contacts were screened to identify program administrators and eligible youth tobacco cessation programs. Those who indicated that they were a program administrator were then asked a series of eligibility questions to determine whether their programs met the operational definition of a youth tobacco cessation programs. The administrators of eligible programs were asked to respond to a survey that would characterize the programs being offered across the United States.

**Analyses**

**Descriptive Statistics.** We report the distribution of contacts by stratification criteria and by sector. Next, we report program screening rates as the proportion of all contacts who were identified as program administrators during the snowball sampling process; eligibility rates were calculated as the proportion of screened programs that met the eligibility criteria; and the program response rate was defined as the proportion of eligible programs that completed the program survey. Screening, eligibility, and response rates were calculated for the sample as a whole and within each sector. In addition, we examined the net contribution of the additional contacts obtained via the Tier 2 sampling, relative to the number obtained from the initial Tier 1 sampling.

**Multivariate Analyses.** We estimated logistic regression models to investigate the degree to which program screening, eligibility, and response rates varied across sectors, after adjusting for initial sector of the contact
and the four stratification variables (urbanization, SES, smoking prevalence, and per capita state tobacco control expenditures).

We used cumulative logit models to analyze the relationship between the number of programs screened, eligible, or profiled in a county (measured as 0, 1, 2, 3, and 4 or greater) and the number of Tier 1 contacts in the county, controlling for the stratification variables and adjusting for state-level clustering. All analyses were conducted with SAS version 9.1.

**Qualitative Information.** Interviewers recorded field notes, which included both quantitative and qualitative information on the degree of difficulty of reaching individual contacts, the challenges encountered across various organizations, and general impressions about the interview process.

### Results

**Sample Yield**

A total of 5,978 Tier 1 key informants were identified within the 408 sampled U.S. counties. Of these nearly 6,000 contacts, 426 (7.1%) were duplicate records, resulting in 5,552 unique Tier 1 key informants of whom 568 (10.23%) were program administrators. The number of Tier 1 informants per county ranged from 1 to 123, with a mean of 10.25 and median 6.00. The snowball sampling yielded an additional 8,107 Tier 2 contacts, of which 3,502 (43.2%) were duplicate records. Among the 4,605 unique Tier 2 contacts, 779 were program administrators, and there were 174 contacts that could not be traced back to the initial sector of contact, due to data entry errors. Therefore, the whole snowball process yielded 9,983 unique and traceable contacts (5,552 unique Tier 1 contacts, 3,826 unique Tier 2 snowball contacts, 779 additional program administrators, and 174 nontraceable records).

**County Stratification and Contact Identification.** By design, the majority of the 9,983 unique and traceable contacts came from counties with high tobacco control expenditures (46%), followed by medium- and low-expenditure counties (30% and 23%, respectively). Compared to low smoking prevalence counties, more contacts were identified in high-prevalence counties (52%); compared with non-MSA counties, more were identified in MSA counties (63%); and more contacts were identified in high SES counties (82%), relative to low SES counties. Across each stratification characteristic, significantly more contacts were identified in Tier 1.
Organizational Sectors and Contact Identification. In the pooled group of Tiers 1 and 2 informants, most contacts were within the education (n = 5,612; 56.2%) and voluntary (n = 2,635; 26.4%) sectors, with smaller numbers identified within the health (n = 825; 8.3%) and other sectors (n = 911; 9.1%). Within the education and other sectors, the majority of contacts were Tier 1 informants (55.6% and 69.9%, respectively), originally identified prior to the snowball process; within the voluntary and health sectors, most contacts were Tier 2 informants, found through the snowball process (54.2% and 54.7%, respectively). All of the noted differences were statistically significant at the .01 level.

Of the total unique and traceable Tiers 1 and 2 informants contacted, 12.4% were screened as possible program administrators (n = 1,236). Program screening rates varied across sectors, ranging from 19.9% and 17.8% of all interviews initiated in the health and other sectors, respectively, to 9.8% and 11.6% of the interviews initiated in the voluntary and education sectors. Of the programs that were successfully screened, 59.1% met the eligibility criteria; eligibility rates varied by sector, with 47.6% from the health sector, 55% from the voluntary sector, 59.5% of the education sector, and over 75% of the other sector contacts qualifying as eligible programs (as would be expected based on the Tier 1 key informant identification protocol). Among the 730 eligible programs, 77.8% completed the program survey; the highest response rate was in the health sector (87.2%), followed by voluntary (82.4%), education (76.5), and other sector (70.5%).

Contribution of Tiers 1 and 2 to Contact Identification. Among the 1,236 unique and traceable contacts screened for eligibility, 524 (42.4%) were from the Tier 1 contacts and the remaining 712 (57.6%) were identified through the Tier 2 snowball process. Of the 730 unique and traceable programs that met the eligibility criteria, 316 (43.3%) were identified at the Tier 1 phase and 414 (56.7%) were found through the Tier 2 snowball process. Finally, of the 568 unique and traceable programs profiled, 230 (40.5%) were identified at the Tier 1 level and 338 (59.5%) through Tier 2 snowball sampling.

Multivariate Models of Screening, Eligibility, and Program Response Rates
Table 1 shows that, in multivariate models that included both sector and stratification criteria as independent variables, there were significantly more screened contacts and eligible programs in urban (MSA) counties,
compared to nonurban counties, but the number of profiled programs was not significantly associated with any of the stratification criteria. Compared to low SES counties, high SES counties yielded more screened program administrators.

Controlling for county stratification characteristics, compared to the education sector, the health sector yielded significantly more contacts that were screened as possible program administrators, significantly fewer programs that fulfilled the eligibility criteria, but nearly twice as many profiled programs. Compared to the education sector, the other sector yielded

| Table 1. Logistic Regression for Program Hit/Eligibility/Response Rates Controlling for Sector and Strata Variables |
|-----------------------------------------------|------------------------------------------------|-----------------------------------------------|
|                                             | Program Screening, N = 1,236 | Program Eligibility, N = 730 | Program Response, N = 568 |
|                                             | OR (95% CI) for OR | OR (95% CI) for OR | OR (95% CI) for OR |
| Sector                                       |                     |                     |                     |
| Health                                       | 2.031 1.674, 2.464   | 0.654 0.460, 0.930   | 2.097 1.032, 4.262   |
| Voluntaries                                  | 0.879 0.752, 1.026   | 0.833 0.620, 1.119   | 1.455 0.887, 2.387   |
| Other                                        | 1.593 1.314, 1.932   | 1.917 1.284, 2.862   | 0.708 0.442, 1.135   |
| Education                                    |                     |                     |                     |
| Tobacco expenditure                          |                     |                     |                     |
| High                                         | 1.128 0.962, 1.322   | 1.270 0.938, 1.720   | 0.659 0.764, 1.764   |
| Medium                                       | 1.109 0.935, 1.315   | 1.343 0.969, 1.861   | 0.694 1.993          |
| Low (reference)                              | 0.996 0.881, 1.125   | 0.919 0.725, 1.164   | 0.638 1.329          |
| Smoking prevalence                           |                     |                     |                     |
| High                                         | 0.996 0.881, 1.125   | 0.919 0.725, 1.164   | 0.638 1.329          |
| Low (reference)                              |                     |                     |                     |
| MSA status                                   |                     |                     |                     |
| MSA                                          | 1.146 1.001, 1.311   | 1.604 1.240, 2.073   | 0.674 1.553          |
| Non-MSA (reference)                          |                     |                     |                     |
| SES status                                   |                     |                     |                     |
| High SES                                     | 1.270 1.061, 1.520   | 0.750 0.527, 1.066   | 0.673 1.975          |
| Low SES (reference)                          |                     |                     |                     |
| Psuedo $R^2$                                  | .0130               | .0280               | .0144               |

Notes: CI = confidence interval; MSA = metropolitan statistical area; OR = odds ratio; SES = socioeconomic status.
significantly more screened contacts and eligible programs but statistically equivalent yields of profiled programs. Table 2 presents the results from our analyses of the relationship between the number of Tier 1 informants in a county and screening, eligibility, and survey response rates. We found that the number of Tier 1 informants was positively and significantly associated with the number of programs screened, eligible, and profiled in that county, controlling for stratification characteristics. In addition, the size of the youth population in a county was also positively and significantly associated with screening, eligibility, and survey response rates. Counties with high per capita tobacco control expenditures were also more likely to have higher rates of eligible and profiled programs, compared to counties with low per capita tobacco control expenditures.

**Qualitative Findings**

Our field notes provide some important context for understanding the snowball sampling process. For example, the key informant and snowball sampling processes were affected by a number of factors, including organization type and region of the county. In terms of organization type, interviewers reported encountering more difficulties in educational organizations than in either voluntary or health organizations. Schools and school districts often had central switchboards that made it difficult to establish contact with someone in the organization, who was knowledgeable about youth tobacco cessation. School districts were also structured differently in different parts of the county; so for example, the process for navigating school districts in New York did not necessarily work while trying to navigate districts in California. With schools subject to academic calendars, conducting snowball sampling was not possible during the summer months and spring and winter breaks.

The region of the country in which key informant and snowball sampling were being conducted also affected the relative ease with which interviewers were able to gather new contacts. Interviewers reported that they had a more difficult time gaining cooperation in regions of the country in which tobacco is a major crop.

**Discussion**

Our goal was to construct a national sampling frame of youth tobacco cessation programs. Our strategy was to conduct snowball sampling to
<table>
<thead>
<tr>
<th></th>
<th>Screened Odds Ratio (p Value)</th>
<th>Eligible Odds Ratio (p Value)</th>
<th>Profiled Odds Ratio (p Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of KIs per county (log scale)</td>
<td>4.23 (&lt;.0001)</td>
<td>3.01 (&lt;.0001)</td>
<td>2.10 (&lt;.0001)</td>
</tr>
<tr>
<td>Youth population (log scale)</td>
<td>1.41 (.034)</td>
<td>1.83 (.002)</td>
<td>1.87 (&lt;.0001)</td>
</tr>
<tr>
<td>MSA (county level) MSA vs. non-MSA</td>
<td>1.32 (.324)</td>
<td>1.30 (.433)</td>
<td>1.42 (.254)</td>
</tr>
<tr>
<td>SES (county level) High SES vs. low</td>
<td>1.51 (.102)</td>
<td>1.16 (.514)</td>
<td>1.22 (.305)</td>
</tr>
<tr>
<td>Smoking prevalence (state level) High smoking vs. low</td>
<td>0.97 (.884)</td>
<td>1.10 (.638)</td>
<td>1.20 (.360)</td>
</tr>
<tr>
<td>Tobacco expenditure (state level)</td>
<td>1.38 (.314)</td>
<td><strong>1.95 (.015)</strong></td>
<td><strong>1.98 (.009)</strong></td>
</tr>
<tr>
<td>Med expenditure</td>
<td>1.24 (.513)</td>
<td>1.64 (.117)</td>
<td>1.89 (.053)</td>
</tr>
<tr>
<td>Low expenditure (reference)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Notes: KIs = key informants; MSA = metropolitan statistical area; SES = socioeconomic status.
saturation within 408 randomly selected counties. Through this process, we gathered contact information for individuals who were identified as being involved with youth tobacco cessation by others within the professional networks that are most closely associated with youth tobacco cessation efforts. Our analyses showed that the majority of informants came from the education sector. Given that the vast majority of youth smoking cessation programs we found were school-based (Curry et al. 2007), it makes sense that the education sector was the most fruitful source of contacts overall.

Our multivariate analyses showed that although the education sector, followed by the voluntary sector, produced the most contacts, the contacts provided by the health and other sectors produced a higher proportion of screened program administrators, compared to either education or voluntary sectors. By far, the other sector produced the highest proportion of eligible contacts, likely because that sector comprised a list of trained providers given to us by an organization that disseminates cessation curriculum. Finally, the proportion of completed surveys was highest within the health sector.

In previous work (Curry et al. 2007), we showed that program availability was associated with county-level stratification criteria: urban status and higher SES. Controlling for sector, we observed that urban and high-income counties yielded significantly more screened program administrators; urban counties produced significantly more eligible programs, but there was no significant association between the county characteristics and program response rate. Our multivariate analyses also suggested that there is a positive relationship between the number of Tier 1 informants initially located and the number of programs screened, eligible, and profiled in a county. Our search for Tier 1 informants was uniform across the counties, so it is likely that counties with greater density of related institutions and individuals likely to be knowledgeable about youth smoking cessation also have greater levels of activity in youth cessation, resulting in higher numbers of screened, eligible, and profiled programs.

Each of the critical elements of our strategy contributed significantly to our ability to find and describe youth smoking cessation programs across the United States. Our operational definition of community-based youth smoking cessation programs was an essential first step, without which we could not have described any of the other elements. Furthermore, in the absence of a sample frame, it was critical to find a proxy-sample frame; counties provided a practical fit for our definition of community and allowed us to stratify our sample so that we achieved representation of characteristics we hypothesized would be related to program availability. Our
results suggest that the breadth of our search was warranted, given that each sector produced significant contributions to our final sample and no single sector consistently dominated across the three stages of screening, eligibility, and survey response. Because our sampling plan reached saturation in each locale, we believe the findings are representative of youth tobacco cessation programs in these areas. Thus, a more narrow approach would have reduced our final sample size and also likely altered the characteristics of our final sample in ways that would have made it less representative. Furthermore, our analyses strongly suggest that the Tier 2 snowball sampling process was a valuable expenditure of our resources. At each stage, from screening, to profiled programs, well over half of our yield came from snowball contacts.

Our qualitative findings illustrate the importance of an adaptable and persistent interviewing team. Our interviewers navigated a variety of organization types, encountering different personnel structures across organization types and regional differences within sectors. They addressed logistical challenges, such as variations in the availability of contacts across the academic year. They also encountered varying levels of cooperation, depending in part on region of the country. For example, the interviewers' reports that it was more challenging to conduct interviews in tobacco-growing regions of the United States were supported by the data: counties in which smoking prevalence was highest produced the smallest percentage of snowball sample referrals (41.7% vs. 47.4% in low-prevalence counties).

Possibly, the biggest challenge was managing the duplicate referrals. Our strategy involved sampling to saturation—asking each new informant to name additional informants within their county until the process exhausted itself and no unique names were gathered. The high number of duplicate nominations provides some evidence that our strategy was successful in sampling to saturation. Through this process, a single person was often nominated several times by others in their professional network. This produced a very large number of referrals that needed to be checked for duplicates on an ongoing basis to ensure that we did not call the same person multiple times. Duplicate checking was made more difficult by the fact that individual names were not always spelled consistently, and standard duplicate checking software is generally not equipped to deal with such discrepancies. For large-scale snowball sampling efforts such as this one, investment in specialized duplicate checking software is recommended.

Some limitations of this research must be considered. For example, we suspect that programs with administrators who were embedded in professional networks within the field of tobacco cessation had a greater
probability of being referred than those with administrators who were more isolated. In addition, those programs with administrators who were not part of the professional networks tapped for sampling may not have been referred. This might include programs affiliated with religious organizations or other community-based organizations that were not known to professionals in educational settings, health organizations, or large voluntary organizations, such as the American Lung Association. Thus, although we exhaustively sampled the networks identified via our key informant and snowball sampling strategy, programs not linked to these networks would not have been accessed using this approach. Our lack of knowledge regarding these programs and whether they even exist should be considered an important limitation of our study.

The strategy we used to identify youth tobacco cessation programs could be used to create a sample frame for other nonprofit organizations that are difficult to identify due to a lack of existing directories, lists, or other sources traditionally used to build sample frames. Several transferable lessons emerged from our experience. First, we clearly defined the parameters of the organizations for which we were searching by developing precise eligibility criteria. Second, we identified a workable unit of analysis for our sampling. Using the county as our unit of analysis enabled us to systematically divide up the US into manageable-sized geographic and population-size units, which we could describe and stratify using publicly available data, and from which we could construct a representative sample using a rigorous and replicable sampling strategy. Next, we organized our search by identifying the likely sectors in each county in which the youth smoking cessation programs could be found. This step enabled us to both narrow our search efforts and more easily identify likely informants. By organizing our search by sector, we were able to contact individuals working closely or directly with youth smoking cessation programs and tap into the professional networks of these individuals through the snowball sampling process. Meticulous record keeping at each stage of the search and interviewing process enabled us to trace our profiled programs back through the snowball process and to analyze the productivity and utility of each stage of our methodology. In sum, our broad, multistep process yielded a large sample of programs from a nationally representative sample of U.S. counties, which could then serve as a potential sample for evaluation.

Note

1. For a more detailed description of the programs identified and profiled, please see Curry et al. (2007).
Declaration of Conflicting Interests
The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research and/or authorship of this article: This study was funded by The Robert Wood Johnson Foundation (Grant 04-6472). Support was also provided by the National Cancer Institute and the Centers for Disease Control and Prevention through the US Department of Health and Human Services, National Institutes of Health, and National Cancer Institute (contract 233-02-0087).

References
Bios

Sherry Emery, MBA, PhD, is a senior research scientist at the Institute for Health Research and Policy and a research associate professor in the Department of Economics at the University of Illinois at Chicago. Her research has focused on tobacco control policy, public health advertising, and program evaluation.

Juhgwha Lee, PhD, MPH, is an instructor in the department of preventive medicine at Northwestern University. Her research interests include statistical methods in epidemiology, longitudinal data analysis, and multilevel analysis.

Susan J. Curry, PhD, is the dean of the University of Iowa College of Public Health. She is recognized internationally for her expertise in behavioral science and translation of research findings into health policy.

Tim Johnson, PhD, is the director of the Survey Research Laboratories and professor in the department of public administration at the University of Illinois at Chicago. His main areas of expertise include survey methodology and health behaviors in disadvantaged populations.

Amy K. Sporer, MS, is an assistant director of research management at the Institute for Health Research and Policy at the University of Illinois at Chicago. She was the deputy director of the Helping Young Smokers Quit project.

Robin Mermelstein, PhD, is the director of the Institute for Health Research and Policy and professor of psychology at the University of Illinois at Chicago. Her research interests fall broadly in the area of tobacco use, with studies ranging from longitudinal examinations of the etiology of youth smoking to cessation interventions for adult smokers.

Brian Flay, DPhil, is a professor of public health in the College of Health Sciences at Oregon State University. His research interests include health promotion and disease preventions research, mass media for health promotion and prevention, smoking and drug use prevention, and prevention research methods and theory.

Richard Warnecke, PhD, is a professor emeritus of epidemiology, public administration, and sociology at the University of Illinois at Chicago. His research has focused over the past 25 years on cancer prevention and epidemiology and health disparities as they related to cancer management.