EVALUATION OF A VOLUNTEER SAMPLE IN NASOPHARYNGEAL COLONIZATION SURVEYS FOR *STREPTOCOCCUS PNEUMONIAE* IN RURAL ALASKA

*Dana L. Bruden, Thomas W. Hennessy, Jay C. Butler, Debra A. Hurlburt, Debra J. Parks, Lisa R. Bulkow*

Arctic Investigations Program, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Anchorage, Alaska

Received 12 July 2004, Accepted 9 November 2004

**ABSTRACT**

**Objective.** To compare characteristics of persons in rural northern communities who participated in a study on antimicrobial use and drug-resistant *Streptococcus pneumoniae* (SP) to those who did not participate.

**Study Design.** The original study (1998-2000) was a community-based, controlled intervention trial designed to determine the penicillin susceptibility of nasopharyngeal SP isolates in relation to community-wide use of antibiotics. The study continued after 2000, in a subset of the original communities, to prospectively evaluate the impact of the heptavalent pneumococcal conjugate vaccine on the carriage of SP. The results presented here are an analysis of the first five years of data.

**Methods.** We conducted annual surveys (1998-2002) for nasopharyngeal colonization of SP using a volunteer sample of residents in rural communities. Medical chart reviews for health clinic visitation and antibiotic use were completed for all village residents.

**Results.** Participants were younger (22.8 vs. 28.4 years), had more health clinic utilization (3.3 vs. 2.4 visits) and received more antibiotics (1.0 vs. 0.6 courses) than non-participants. Differences between participants and non-participants were similar across all years of the study.

**Conclusions.** Our study provides further empirical evidence that selection bias should be considered when designing studies. However, a volunteer sample provided internal consistency for comparison of our main study outcomes across time. (*Int J Circumpolar Health* 2005;64(1):16-25.)

**Keywords:** participation, non-participation, non-randomized sample, selection bias, Alaska, community intervention
INTRODUCTION

A random sample, necessary for drawing conclusions about a population based on a study sample, is logistically difficult to obtain, and is cost-prohibitive in medical studies. It is often impossible to obtain a list of all persons with a given medical condition or from a given study population of interest, from which to draw a random sample. Often, studies must rely on a person making contact with the health care system, before they are recruited into a particular study. Additionally, in most studies concerning human subjects, potential participants must consent to take part, introducing selection bias into a sample. Relying on a volunteer sample of subjects that have contacted the health care system, results in a non-representative sample with possible bias in the outcomes of interest.

We assessed the bias associated with a non-representative sample in the context of a study in remote Alaska. The primary outcome measure in the original study was the prevalence of nasopharyngeal carriage of penicillin-nonsusceptible *Streptococcus pneumoniae* (SP) over time. Because we suspected our participants were likely to have used antibiotics more often than non-participants, and since antibiotic use is associated with resistant pneumococci (1-4), the estimate of our primary outcome measure was potentially biased. The original study involved two interventions in 17 villages in rural Alaska (1998-2000) to determine if either education about judicious use of antibiotics (5), or an increase in the routine dose of amoxicillin (for children under 7 years of age) (6), could be successful in decreasing the proportion of persons carrying penicillin-non-susceptible *Streptococcus pneumoniae* (SP). The communities involved are accessible only by boat, or air travel, and range in size from 75 to 800 persons. Health care is provided in each village at a primary care clinic staffed by community-health aides, or physician assistants. Each intervention was developed to target four communities within a region, while the other two regions (nine villages) were left as controls. Treatment was assigned at the regional rather than the individual level. To evaluate whether the educational intervention was successful in reducing antibiotic prescribing, medical charts at each village clinic were reviewed for all residents (5). To assess the impact of the interventions on the carriage of penicillin-nonsusceptible SP, nasopharyngeal swab specimens were collected from a volunteer sample in each of the villages in the first 3 years (1998-2000) of the study (5). The study continued after 2000 in order to evaluate the impact of the heptavalent pneumococcal conjugate vaccine on the carriage of SP.

A random sample of village residents was not considered, as it would have involved a door-to-door recruitment process, which was considered culturally inappropriate. The lack of a random sample precludes generalization of the study results to the entire community based on statistical design and prompted examination of the demographic representativeness of the volunteer sample. Because we conducted chart reviews on all residents of the participating villages, medical and demographic characteristics of the participants could be compared to those who did not participate. Herein, we report on a secondary objective of the project: to examine characteris-
tics of a volunteer sample in rural Alaskan villages, and to evaluate the effect of the volunteer sampling technique on the main study objective.

METHODS

Study design
Originally (1998-2000), the study was conducted in four regions (17 villages) of rural Alaska. The regions were selected on the basis of adequate geographical distance between them to eliminate any crossover effects from one region to another during the initial intervention phase. Villages were selected within the regions based on willingness to participate assessed during community council meetings. No communities that were approached refused participation. Details regarding the community-based interventions and primary outcomes from the initial study have been provided previously (5, 6). After 2000, the study continued in two of the original four regions (8 of the original 17 villages), in order to evaluate the impact of the introduction of the heptavalent pneumococcal conjugate vaccine on community-wide carriage of SP. To assess the impact of the original interventions and of the introduction of the vaccine on SP carriage, annual nasopharyngeal surveys were conducted in each participating village during the months of April and May (1998-2002). For the annual nasopharyngeal survey, study personnel spent 2-4 days in each village and operated out of the community health clinic. Residents were asked to participate in the study through the use of study posters [displayed at the health clinic, community hall and store], radio announcements, and letters sent to each household. Residents were reimbursed $25 to compensate them for the time to participate in the study. Participating in the study involved a brief interview regarding demographics and medical history, and a nasopharyngeal swab. Methods employed for the nasopharyngeal swab specimen handling and laboratory procedures have been described previously (5) and remained consistent during each study year reported in this analysis (1998-2002).

At the community health clinic, an individual medical record is maintained for each village resident by community health aides, or physician assistants. As a secondary outcome in the original study, medical records of all village residents were reviewed to determine the number of clinic visits (excluding preventive health visits) and the baseline number of antibiotic courses prescribed during the previous 6-month period from October 1 to March 31. Waver of individual consent for medical chart review of all residents was obtained at the village council, regional health corporation and the state-wide institutional review board. Methods of the medical chart review have been published previously (5) and have remained consistent throughout the follow-up years involved in this analysis. This study was reviewed and approved by Institutional Review Boards of Centers for Disease Control and Prevention (CDC), Indian Health Service, and the participating Native health corporations.

Statistics
We compared the characteristics (antibiotic use, clinic use, age) of those who participated...
in the study to those who did not using a generalized linear model, with village and region entered into the model as fixed effects. The sampling method was identical in the first three years of the study and provided the opportunity to test if the characteristics of participants and non-participants were similar in the first 3 years of the study. If this were the case, the study would provide internal validity for the comparison of penicillin-nonsusceptible SP rates across years (1998-2000).

We examined characteristics (clinic use and antibiotic use) of participants and non-participants between years by adding year, participation and age into the generalized linear model and testing for an interaction between year and participation (coded as an indicator variable). A statistically significant interaction between year and participation would indicate that the characteristics of the participants and non-participants differed depending on study year. All two-way interactions and the three-way interaction term between region, year and participation were entered into the model. Backwards elimination was used for final model selection using the Wald chi-square test statistic.

For 2001 and 2002, in the eight villages (two regions) that remained in the survey, we compared characteristics of new participants in the study to persons who had participated in a previous year. New participants were defined as village residents participating in the 2001 or 2002 nasopharyngeal swab surveys that had not participated in any survey from 1998-2000. Previous participants were defined as village residents participating in the 2001 or 2002 nasopharyngeal swab surveys who had also participated in at least one of the surveys between 1998 and 2000. Children ≤ 3 years of age for 2001, and ≤ 4 years of age for 2002, were removed from these comparisons, because they were not alive at the time of the 1998 survey and, consequently, were more likely to be new participants. Sex, SP carriage rates and penicillin-nonsusceptible SP rates were compared between new (previous non-participants) and previous participants by the Mantel-Haenszel test, stratified by village of residence. Age, the number of courses of antibiotics and clinic visits were compared as previously described.

A simulation was used to estimate the SP carriage rate and penicillin nonsusceptible SP carriage rate in those village residents who were not sampled. For each resident not sampled, a predicted probability of SP carriage was obtained from a logistic regression model based upon the data of the participants. The first set of logistic regression models (one for each study year) included age class (0-9, 10-14, 15-19, 20-29, 30-54, and 55+ years), study region, and the number of antibiotic courses. The predicted probability was used to classify the non-participant as a carrier, or non-carrier of SP. If carriage of SP was predicted for an individual, a second logistic regression model was used to determine the probability that the carried organism would be non-susceptible or sensitive to penicillin. The second set of logistic regression models included age class (0-1, 2-4, 5-18, 19+), study region and the number of antibiotic courses. The predicted probability from the second logistic regression model was used to classify the individual as carrying penicillin sensitive, or non-susceptible SP. Outcomes from the logistic regressions were summed to obtain predicted SP
and penicillin-nonsusceptible SP carriage rates for the resident non-sampled population. The simulation was run 10,000 times and the mean rates were calculated over all iterations. Bias was then calculated as the difference between the rate from the volunteer sample and the estimated rate from the entire population. Because the objective of the paper is to characterize participants and non-participants in our surveys, rather than to determine risk factors for drug resistant SP carriage as in previous analyses (5), the results from logistic regression models are only presented indirectly, through reporting of the simulation results. All p-values are two-sided and < 0.05 was considered statistically significant.

RESULTS

Participant characteristics (1998 - 2000)
Over the first 3 years of the study, 5,870, 6,255 and 5,937 medical charts were reviewed during 1998-2000, respectively. The participation rates in the nasopharyngeal SP culture survey were 33 % (n = 1,956), 30 % (n = 1,873) and 36 % (n = 2,114) in 1998 - 2000, respectively. In 1998, the mean age of participants was 22.8 years, which was significantly younger than the 28.4 years for non-participant residents (Table I). Compared with non-participants during the previous 6 months, participants were more likely to have visited the clinic, had more clinic visits, were more likely to have been prescribed antibiotics, and were prescribed more courses of antibiotics (Table I). After controlling for age, there was still a significant difference in the number of antibiotic courses (p < 0.01) and clinic visits (p < 0.01) between participants and non-participants in 1998 (figure 1). Participants were significantly younger in age, had more clinic visits and antibiotic courses than non-participants in 1999 and 2000 (Table I), similar to the results for 1998. When we restricted the analysis to persons known to have used the community health clinic during the chart review period (number of clinic visits ≥1), results comparing age, antibiotic and clinic use of participants and non-participants (1998-2000) were similar (data not shown).

Impact of sampling method on initial study objective
To examine differences between participants and non-participants across years, the number of antibiotic courses and clinic visits per person for study participants and non-participants in rural Alaska, 10/1/1991-3/31/1998.
### Table I.
Age, number of clinic visits and antibiotic courses for rural Alaskan participants and non-participants in an annual survey for SP colonization, Alaska, 1998-2000.

| Characteristic, or outcome | Year | Participants | Non-participants |
|----------------------------|------|--------------|------------------|
| Mean age, years (% ≤ 20 years of age) | 1998 | 22.8 (55%) | 28.4 (38%) * |
|                              | 1999 | 22.5 (56%) | 28.5 (38%) * |
|                              | 2000 | 23.8 (54%) | 29.2 (36%) * |
| Mean number of clinic visits† | 1998 | 3.31 (81%) | 2.38 (67%) ** |
| (% with ≥ 1 clinic visit)    | 1999 | 3.29 (80%) | 2.20 (66%) ** |
|                              | 2000 | 2.87 (78%) | 2.01 (64%) ** |
| Mean number of antibiotic Courses† | 1998 | 1.02 (49%) | 0.65 (36%) ** |
| (% with ≥ 1 antibiotic course) | 1999 | 0.79 (44%) | 0.53 (32%) ** |
|                              | 2000 | 0.72 (42%) | 0.46 (28%) ** |

* P-value comparing mean for participants and non-participants < 0.05, stratified by village.
** P-value comparing mean for participants and non-participants < 0.05, stratified by village and age group.
† Mean number of clinic visits and antibiotic courses for the 6-month period (Oct. 1 - Mar. 21).

### Table II.
Comparison of characteristics of participants by previous participant status, 2001-2002.

| Outcome measure | Participants in 2001 (n = 2,463)† | Participants in 2002 (n = 2,066)‡ |
|-----------------|-----------------------------------|-----------------------------------|
| Mean age (years) | 25.3 (771/1,696) | 29.4* (461/767) |
| Mean number of antibiotic courses | 0.63 (471/771) | 0.52** (225/461) |
| Mean number of clinic visits | 3.00 (461/1,696) | 2.20** (225/771) |
| Sex (% male) | 45.5% (771/1,696) | 60.1%*** (461/767) |
| % Carrying S. pneumoniae | 38.1% (461/1,696) | 29.3** (225/771) |
| % Penicillin non-susceptible | 21.8% (461/1,696) | 18.7% (225/771) |

* P-value comparing new participants and those who had previously participated < 0.05, stratified by village.
** P-value comparing new participants and those who had previously participated < 0.05, stratified by village and age group.
† Limited to those ≥ 3 years of age.
‡ Limited to those ≥ 4 years of age.
§ Participants in 2002 who had previously participated in 2001 only (n = 465) not included [Mean age = 30.9 years, antibiotic courses = 0.51, clinic visits = 2.27, male = 59.6 %, SP carriers = 29.3 %, PNSP = 16.9 %].
of clinic visits and antibiotic courses were entered into a linear model with year, treatment region, village, age and participation. The interaction term between participation and year was not significant in either the model for clinic visits (p = 0.42), or the number of antibiotic courses (p = 0.25), indicating that the differences between participants and non-participants were similar in the first 3 years of the study. For example, participants were prescribed 57%, 49% and 57% more antibiotics than non-participants in 1998, 1999 and 2000, respectively. Participants had 39%, 49%, and 43% more clinic visits than non-participants in 1998, 1999 and 2000, respectively. We found the increased antibiotic and clinic use among participants to be consistent across all regions of the study area (data not shown).

Newly recruited participant characteristics (2001-2002)
Participation rates increased in 2001 (63 %, n = 2,662) and 2002 (62 %, n = 2,260) from the overall baseline participation rate of 1998-2000 (33 % in all four regions and 38 % in two regions that continued to participate in 2001/2002). The increase in participation may have been due to a change in reimbursement procedures (check provided at the time of participation, rather than mailed) and an increased familiarity of village residents with the study procedures and objectives. We compared characteristics of participants who were newly recruited to the study in 2001 and 2002 (previous non-participants in 1998-2000), to those who had participated in the study during the first 3 baseline years. In both 2001 and 2002, new participants (previous non-participants) were older in age, had been prescribed fewer antibiotics and had fewer clinic visits than participants who had enrolled in at least 1 year between 1998 and 2000 (Table II). In both years, new participants (previous non-participants) were more likely to be male (60.1 % male, 2001; 63.2 % male, 2002) than previous participants (45.5 % male, 2001; 45.4 % male, 2002). This difference in gender between new (previous non-participants) and previous participants was less evident for children under 10 years of age than among adults and adolescents aged 10 years, or more (data not shown). In 2001, new participants (previous non-participants) were less likely to carry SP (29.3 %) than those who had also participated in one of the baseline years (38.1 %, Table II). However, this difference was not repeated in 2002 (new participants [34.3 %], previous participants [31.1 %]). In both years, there was no difference between the newly recruited participants (previous non-participants) and those that had participated during the baseline years, in the proportion of carriers with penicillin-nonsusceptible SP (Table II).

Predicted carriage and resistance prevalence among non-participants
Predicted probabilities from two logistic regression models were used to simulate carriage and resistance for SP in non-participants from the 5 study years (Table III). The estimated SP carriage rate in non-participants was consistently lower than the actual carriage rate in participants. The difference between the actual rate in participants and the estimated rate for non-participants was highest in 1999 (30.4 % vs. 22.9 %) and lowest in 2002 (35.1 % vs. 31.6 %). The estimated bias
Table III.
Carriage of *S. pneumoniae* and the percentage of penicillin-nonsusceptible *S. pneumoniae* isolated for those that participated in the volunteer sample with the estimated SP carriage rate, and percentage penicillin-nonsusceptible for those who did not participate in the volunteer sample, and the estimated bias, rural Alaska, 1998-2002.

| Study Year | Actual for participants | Estimated rate for non-participants | Estimated rate for population | Bias | Actual for participants | Estimated rate for non-participants | Estimated rate for population | Bias |
|------------|-------------------------|-------------------------------------|-----------------------------|------|-------------------------|-------------------------------------|-----------------------------|------|
| 1998       | 33.0%                   | 26.1%                               | 28.4%                       | 4.6% | 28.4%                   | 26.0%                               | 26.9%                       | 1.5% |
| 1999       | 30.4%                   | 22.9%                               | 25.1%                       | 5.3% | 24.6%                   | 23.2%                               | 23.7%                       | 0.9% |
| 2000       | 32.2%                   | 25.6%                               | 28.0%                       | 4.2% | 31.1%                   | 29.3%                               | 30.0%                       | 1.0% |
| 2001       | 36.3%                   | 31.9%                               | 34.7%                       | 1.6% | 22.2%                   | 20.8%                               | 21.4%                       | 0.8% |
| 2002       | 35.1%                   | 31.6%                               | 33.7%                       | 1.3% | 17.6%                   | 16.9%                               | 17.4%                       | 0.3% |

in the SP carriage rate ranged from 1.3% to 5.3%. The actual penicillin-nonsusceptible SP rate for participants was only slightly higher than the estimated penicillin-nonsusceptible SP rate for non-participants. The greatest difference between the two rates was observed in 1998 (28.4% vs. 26.0%) and the smallest, in 2002 (17.6% vs. 16.9%). The estimated bias in the penicillin-nonsusceptible SP rate ranged from 0.3% to 1.5%.

DISCUSSION

In annual non-randomized surveys of SP colonization in rural Alaska, we found that our participants were younger in age, had more visits to the health care clinic and had been prescribed more antibiotic courses than the non-participants. In our study, we found a higher level of antibiotic use in participants, which is a well-known risk factor for carriage of penicillin-nonsusceptible SP (1-4). Although differing in survey design and instrument, some studies have found a higher prevalence of risk factors in participants of health surveys, as was the case with cardiovascular disease (7), while other studies have found lower levels of risk factors in participants, as was the case in studies related to life expectancy and smoking (8, 9). However, the characteristics of our participants and non-participants were consistent across years, which allowed for internal comparison of our outcome of interest (% penicillin-nonsusceptible SP) over time. We are not aware of any other studies that compare antibiotic prescribing rates, or rural health clinic utilization, by volunteer participation in a health-related survey.

A large increase in the participation rate in a subset of the original study regions between the baseline years (1998-2000) and 2001 and 2002 allowed us to examine characteristics of a proportion of the population we had not sampled for SP colonization during the first 3 years of the study. The characteristics of these new participants (previous non-participants) followed what would have been predicted by differences between participants and non-participants in the first 3 years of the study, in that they were older in age and had less antibiotic and health clinic usage than those community members that had also participated in the first 3 years of the study. We
found that new participants (previous non-participants) were more likely to be male than those we had seen in previous surveys, which is in agreement with findings from other volunteer surveys (10-13). New participants (previous non-participants) were less likely to carry SP than those that had previously participated in the surveys. However, this difference diminished in 2002 and disappeared in 2001 after stratification by age group.

Studies that rely on participants that have some form of contact with a health care system, while not being representative of a general population, may be representative of the population of health care seekers. In our study, restricted to the population in a rural setting that had sought health care at their village clinic during a 6-month chart review period, volunteer participants still had higher health clinic utilization and a higher antibiotic prescription rate than non-participants. Therefore, participants were not representative of the overall population of health care users.

Our study had several limitations. The validity of the estimated SP and penicillin-nonsusceptible SP carriage rates were based upon the assumption that the relationships between those outcomes and age, study region and antibiotic courses, were the same for non-participants as those observed amongst participants. This assumption is not testable. We considered all persons with a medical chart in the village clinic to be village residents and classified them as non-participants if they did not take part in the survey. Village clinics maintain, in a separate location, medical charts of persons known to have moved from the village, and these charts were not reviewed. Because of the small size of the villages and familiarity of clinic staff with residents, we believe this is done accurately. However, to eliminate this potential misclassification, we examined differences between participants and non-participants among persons known to have lived in the village during the chart review period (≥ 1 health clinic visits) and results were the same. The new heptavalent conjugate pneumococcal vaccine was administered to many children under the age of five in these communities in 2001 and 2002. It has most likely contributed to the decline in the age of persons carrying penicillin-nonsusceptible SP observed in recent years among participants (Table III). We did not collect data on the administration of this vaccine in non-participants and, consequently, it could not be included in the carriage simulations for those years.

We believe the information presented here will be useful to other researchers conducting studies in similar rural settings. Our principal finding, that volunteer study participants are younger in age, more likely to be female, to have sought health care, and to have been prescribed antibiotics, highlights the practical importance of valid sampling techniques whenever they are possible in medical research. In our setting, where random sampling was not feasible, a volunteer sample provided internal consistency for comparison of our main study outcome across time.

Acknowledgments
This study was supported in part by the National Center for Infectious Diseases' Prevention Working Group and the National Vaccine Program Office at the Centers for Disease Control and Prevention, Atlanta, GA. We thank Helen Peters, James Gove and Catherine
Dentinger for their leadership in collecting data on field trips; Carolyn Zanis, Carolynn DeByle, Dr. Karen Rudolph and Karen Miernyk for their expertise in handling laboratory specimens in the field; Alisa Reasonover, Marcella Harker-Jones and Julie Morris for their long hours identifying *Streptococcus pneumoniae* with drug resistance; and Terri Brandon and Pat Bering for administrative support.

REFERENCES

1. Austin DJ, Kristinnson KG, Anderson RM. The relationship between the volume of antimicrobial consumption in human communities and the frequency of resistance. Proc Natl Sci USA 1999;96:1152-1156.
2. Baquero F, Martinez-Beltran J, Loza E. A review of antibiotic resistance patterns of *Streptococcus pneumoniae* in Europe. J Antimicrob Chemother 1991;28:C:31-38.
3. Leach AJ, Shelby-James TM, Mayo M, et al. A prospective study of the impact of community-based azithromycin treatment of trachoma on carriage and resistance of *Streptococcus pneumoniae*. Clin Infect Dis 1997;24:356-62.
4. Criqui MH, Barrett-Conner E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. Am J Epidemiol 1978;108:367-372.
5. Hennessy TW, Petersen KM, Bruden D, et al. Changes in antibiotic prescribing practice and carriage of penicillin-resistant *Streptococcus pneumoniae*: a controlled intervention trial in rural Alaska. Clin Infect Dis 2002;34:1453-1450.
6. Hennessy TW, Bruden D, Petersen KM, et al. Effect of high-dose amoxicillin on the prevalence of penicillin-resistant *Streptococcus pneumoniae* in rural Alaska. JAMA 2002;287(16):2078-2079.
7. Doll R, Hill AB. Mortality in relation to smoking: ten years’ observations of British doctors. BMJ 1964;1:1399-1410.
8. Hammon EC. Life expectancy of American men in relation to their smoking habits. J Natl Cancer Inst 1969;43:951-962.
9. Hansen V, Jacobsen BK, Arenesen A. Prevalence of serious psychiatric morbidity in attenders and non-attenders to a health survey of a general population. Am J Epidemiol 2001;154(10):891-894.
10. Vernon WV, Roberts RE, Lee ES. Ethnic status and participation in longitudinal health surveys. Am J Epidemiol 1984;119(1):99-113.
11. Comstick GW, Helsing KT. Characteristics of respondents and non-respondents to a questionnaire for estimating community mood. Am J Epidemiol 1973;97:233-239.
12. Clark VA, Aneshensel CS, Frerichs RR, et al. Analysis of non-response in a prospective study of depression in Los Angeles County. Int J Epidemiol 1983;12:193-198.

Dana Bruden, MSC
Arctic Investigations Program/CDC
4055 Tudor Centre Dr.
Anchorage, AK 99508-5932
Email: zk9@cdc.gov