

SAMPLING MEN WHO HAVE SEX WITH MEN: STRATEGIES FOR A TELEPHONE SURVEY IN URBAN AREAS IN THE UNITED STATES

D. Binson, J. Moskowitz, T. Mills, K. Anderson, J. Paul, R. Stall, and J. Catania
University of California, San Francisco

Diane Binson, Center for AIDS Prevention Studies, University of California, 74 New
Montgomery, San Francisco, CA 94105

Key Words: Sampling, Gay Men, AIDS

Introduction

Men who have sex with men (MSM) have become a major focus in research on sexuality in the context of AIDS. Estimates from recent national probability surveys indicate that approximately 2% to 3% of men report sexual contact with another man during the year prior to being interviewed (Fay, Turner, Klassen, & Gagnon, 1989; Rogers and Turner, 1991; Catania, Coates, Stall, et al., 1992; Laumann, Gagnon, Michael, and Michaels, 1994; Binson, Michaels, Stall, et al., 1995). One limit to population based surveys conducted to date, however, has been the small number of gay or bisexual men or MSMs in these samples. These small numbers preclude more complex analyses of the psychosocial correlates of same sex behavior among men, especially minority men. Only in general population surveys of at least 20,000 interviews would there be adequate numbers of gay and bisexual men for analysis. Efforts to study men who have sex with men utilizing larger samples, for the most part, have depended on convenience samples drawn from clinics, organizational memberships, and other kinds of listings. An exception to this have been several large studies based on probability samples in Los Angeles and San Francisco (Stall, et al., 1992; Kanouse, et al., 1991; Winkelstein, 1987; Osmond, et al., 1994). Except for the study in Los Angeles and one telephone survey in San Francisco (Stall, et al., 1992), all studies have been face-to-face surveys. While San Francisco and Los Angeles are important epicenters, they only account for 17% of AIDS cases among men who have sex with men. No studies have been conducted of men who have sex with men utilizing probability sampling techniques in other areas of the United States. This is the case, despite the fact, that men who have sex with men account for a majority of all AIDS cases (CDC, 1995, 7, no. 2) and over 80% of these cases are in cities outside of San Francisco and Los Angeles.

In the *Multicultural Men's Health Study (MMHS)*; supported by grant MH54320 from the National Institute of Mental Health), we have proposed to conduct interviews with 5,500 gay and bisexual men, and men who have sex with men who do not identify as

gay or bisexual. The content of the survey will focus on five general topic areas: (a) sexuality and sexual networks; (b) HIV-related risk behaviors and testing; (c) physical and mental health; (d) violence and discrimination; and (e) caregiving and bereavement. Data collection will begin in late 1996 in four cities (New York, Los Angeles, Chicago, and San Francisco). Samples will be drawn utilizing probability sampling techniques with an oversample of African American and Latino men. Interviewing has been proposed in six additional cities (Boston, Atlanta, Miami, Dallas, San Diego, and Washington, D.C.) and interviewing in these cities is contingent on the cost of conducting interviews in the first four cities. The population is men residing in targeted neighborhoods in these cities, who self-identify as gay or bisexual or who report having had sex with men at any time since the age of 14 but may not identify as gay or bisexual. Although the actual probability of selection and the total number of gay/bisexual/MSM who reside in particular geographical areas are unknown, using RDD procedures will yield a probability sample, since each self-identified (i.e., willing to identify themselves to an interview over the phone) gay/bisexual/MSM residing in the targeted geographic area has an equal probability of selection. In each sampled area of the city, telephone households will be selected using a combination of RDD (Waksberg, 1978) and list-assisted frames. A strict Mitofsky-Waksberg design could be prohibitively expensive in some cities. The combination of list-assisted and RDD methods will be used to test the efficiency of the Mitofsky-Waksberg design versus the list-assisted method in geographical areas in which the targeted population density appears to be high.

The sample design for this study involved three major stages. The first was to locate general areas of the city in which eligible respondents resided. The second, was to define precisely the geographical boundaries of the targeted area, which in this study were designated by zip code boundaries. The third stage, which is still in progress, involves selecting exchanges (within each zip code) and designating the exchanges to be included in the sample for each city. This paper will describe stage one and two and will include a brief overview of the issues involved in stage three. A more detailed discussion of stages two and three will be described in a later paper.

Representativeness versus Costs

Our goal in the *MMHS* is to draw a sample that will be representative of the widest possible population of gay or bisexual men or MSM residing in each of the selected cities. Our intention is to be as inclusive as possible. We do not want to define the eligible population only in terms of behavior, as has been done in many AIDS-related studies, nor do we want to define the eligible population only in terms of identity. Instead, we want to include men who consider themselves to be gay or bisexual but are not necessarily having sex or not necessarily having sex with men; and we want to include men who have sex with men but do not necessarily identify themselves as gay or bisexual. In terms of a sampling strategy, the challenge was clear: how could we include the widest possible areas of the city, without having to use most of our money screening households only to find an extraordinarily small number of eligible respondents? In large cities in the US gay/bisexual/MSM live in many areas of the city, but gay men and also probably bisexual men are more heavily concentrated in some areas. However, we didn't want to confine ourselves only to "gay neighborhoods," as we might miss large numbers of the target population who live outside those neighborhoods. We wanted to find a way to identify concentrations of the target population beyond the "gay neighborhoods." Unfortunately there are no reliable estimates of the proportion of the population in any given city who are either gay or bisexual men or MSM, nor is the residential distribution of this population known. In order to identify higher concentrations of the target population beyond just the "gay neighborhoods", we used five measures and a mapping program, to translate each of the measures to a graphical representation.

Stage One: Identifying Residential Clusters

We identified the locations of residential clusters of gay/bisexual/MSM in the 10 cities using five sources of information: (1) US. Census data on same gender partnered households; (2) cumulative AIDS case load information by zip code; (3) marketing lists designed to reach gay men; (4) gay related businesses and services; and (5) local informants. In addition, each city was visited by a team of researchers to establish local contacts, explain the purposes of the study, obtain feedback, enlist the support of the local community, and confirm with knowledgeable city residents the residential clusters of gay and bisexual men that we had identified. The five measures are described below:

1. US Census. In the 1990 census a question was asked that identified same gender partnered households. Adults were asked to specify their

relationship to other adults in the household. If they were not related--as a husband or wife, for example--they could indicate that they were "unmarried partners" as indicated below. The Census Bureau provided us with these data for male partnered households by block for each of the cities. One problem with these data are that they are somewhat out of date and population shifts within cities and to the suburbs are not represented by these numbers. There is also a political issue that may have led to some variation in responses. In some cities, for example, men were encouraged by the gay community to "stand up and be counted", while in others there was less enthusiasm to provide this kind of information to a government agency. In addition, this measure represents only one subgroup--partnered men who cohabit with men. It doesn't count men who are partnered but do not cohabit with their partner, nor does it count men who are unpartnered. It is also a more sensitive measure to men who are "out" and who are willing to disclose their identity to a stranger.

Question Identifying an Unmarried Partner in 1990 Census

For other adults in household:

How is this person related to PERSON 1?

Husband/Wife
Natural-born or adopted son/daughter
Stepson/stepdaughter
Brother/sister
Father/mother
Grandchild
Other relative (Print exact relationship)

If NOT RELATED to person 1:

Roomer, boarder, or foster child
Housemate, roommate
Unmarried partner
Other nonrelative

2. Cumulative AIDS cases (through 1994): AIDS cases were compiled from local health departments. We counted two categories of cases, those that were attributed to sex between men and those cases attributed to men who had sex with men and who were also intravenous drug users. The data were by zip code and three race and ethnic categories (white, African American, and Latino). In actuality, however, this measure more likely reflects where the epidemic has been than it does areas of the city in which the targeted population resides. In addition, one could say that AIDS cases represent sex with infected men and, as such, would reflect different patterns in each city

depending on the prevalence and history of HIV infection. To help identify geographical areas that would facilitate drawing an oversample of African American and Latino residents in each city, we used AIDS cases (in the same two categories described above) among African American and Latino men. If the proportion of African American and Latino AIDS cases within a particular zip codes (number of cases divided by number of households) was higher than 1%, we designated that zip code as an African American or Latino zip code.

3. **Marketing Lists:** A third measure was a list of addresses compiled from a variety of commercial marketing lists. These lists are of contributors to gay/lesbian and AIDS organizations, contributors to political campaigns for gay/lesbian candidates, consumers ordering from gay catalogs, subscribers to gay magazines, and bar lists. There were a total of 500,000 names on the list, 300,000 of whom were men, 100,000 were women, and 100,000 were unknown. The list is updated continuously with address changes and then checked quarterly. We used addresses identified as belonging to men in each of the selected cities (68,000). It is reasonable to assume that these kinds of marketing lists probably are more likely to contain names of those who are "out" and those involved in the gay community. We found in the pretest we did using telephone numbers from these lists that the lists contained mainly men who were disproportionately older, white, with high incomes, and high levels of education.

4. **Gay-related Businesses and Services:** The Gayellow Pages is a guide to accommodations, AIDS resources, bars businesses, churches, healthcare, lawyers, organizations, publications, across the US. We compiled addresses from the Gayellow Pages of the following types of businesses/services: AIDS/HIV support, education, advocacy, and publication; bars, restaurants, clubs/discos, and Gay bookstores; erotica; organizations/resources (from business and professional to social and support groups). One drawback of the list of service addresses from the Gayellow pages is that businesses pay to advertise in it so the addresses are a select group of the services for gay men.

5. **Local Informants:** We contacted at least three knowledgeable informants in each city for their opinions of street boundaries of gay neighborhoods. We emphasized that we were looking for areas in which gay men live in higher concentrations than the rest of the city, not just the areas with gay bars and businesses.

Given that none of these pieces of information was perfect, we looked for general agreement among the sources to identify geographical areas of the city containing the eligible population. This information

was mapped using ArcView 2.0, a computer mapping program that allowed us to produce transparency overlays for each city so we could identify where the sources of information overlap. When the individual maps for each measure were superimposed over each other, they confirmed a general concentration of the target population in similar areas of the city. The map shown on the next page for San Francisco (see Figure 1) displays three maps representing three measures-- AIDS cases, census data, and marketing list addresses-- superimposed on each other. For San Francisco, and most other cities, the map representing the overlay of individual maps indicated that the target population is located in a much larger geographical area than was indicated by any one of the measures viewed separately.

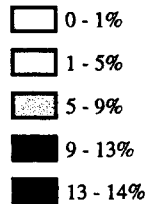
Stage Two: From Locating Residential Clusters to Sampling Zip Codes

Once we identified general areas of the city where concentrations of the eligible population resided using these multiple measures, we had to "translate" the information in such a way that the targeted area had clear and exact boundaries. One possible approach was to use block group data since the census data were at the block group level. However, using block level data was problematic in that selected block groups based on a large proportion of same gendered partner households (using the census measure) were not always contiguous. This meant that describing the sample as representing the city population living in recognizable areas of the city would have been cumbersome. Hence, the decision was made to use zip codes as the "neighborhood" boundaries.

The purpose of stage two was to estimate the prevalence of MSMs by zip code. This information was needed in order to determine which zip codes indicated in stage one should be kept and included in the final sample. We did this in three steps. The first step was to estimate the total number of MSMs in the selected cities. The second step was to estimate the number of MSMs in each city. The third step was to estimate how many MSMs were in each zip code.

For step one, to calculate the total number of MSMs in the selected cities we used two pieces of information: the number of males 18 years or older from the census and an estimated prevalence of MSMs one would expect in these cities from Binson, et al. (1995). For step two, we used four types of data: AIDS incidence adjusted for HIV prevalence, marketing lists, 1990 census data, and Holmberg's (1996) estimates of MSMs in metropolitan statistical areas. We distributed the number of MSMs in each city so that they would be proportional to these four types of data.

Census data: % Unmarried male partner households



MSM AIDS Cases/Households (10 + %)

MSM AIDS Cases/Households (5 - 10%)

MSM AIDS Cases/Households (1 - 5%)

Marketing List

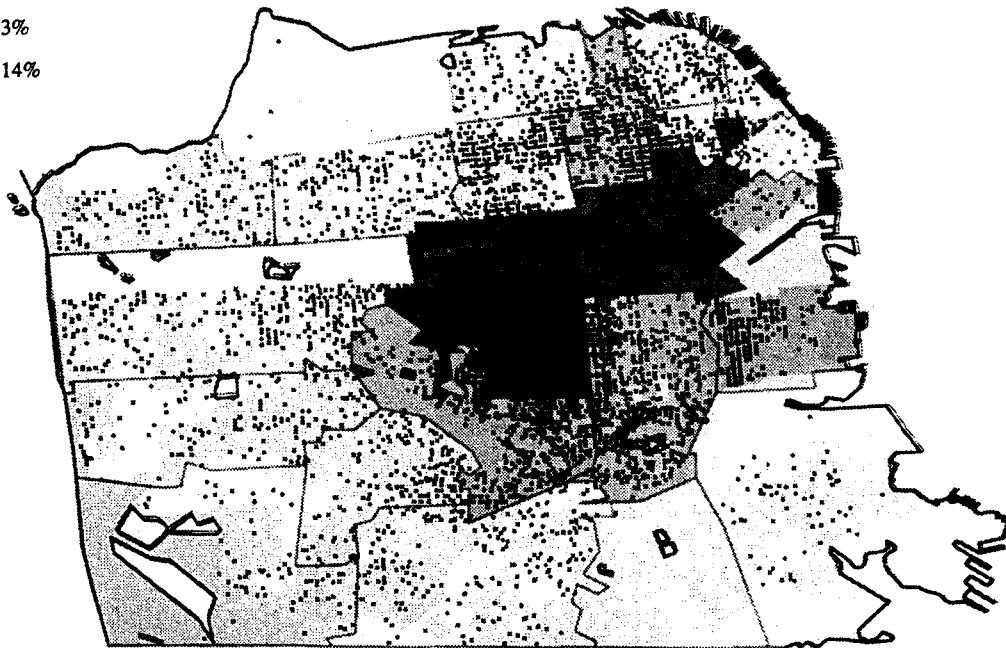


Figure 1: San Francisco

The four estimates for each city were then averaged so that we would have a single estimate of MSMs for each city. The third step was to estimate the distribution of MSMs in each city by zip code. Of these four types of data, the only one we had by zip code was the marketing list. We distributed the MSMs among the zip codes so they would be proportional to the number of men on the marketing list. Only zip codes with a minimum prevalence (yet to be determined) will be included in the sample for each city.

Stage Three: From Zip Codes to Exchanges

In stage three, Genesys Sampling Systems provided a list of exchanges for each zip code, by the proportion of the zip code covered by a particular exchange and by the proportion of the exchange in the zip code. With this information it is possible to calculate the expected cost per interview in each exchange. Based on this information, we will determine which exchanges would be too costly to be included given our resources and they will be deleted from the final sample frame.

Conclusion

In situations where sampling the entire city is not possible, it is helpful to locate a good part of the target population using multiple measures. Because each of the different measures was on a different metric (zip codes, addresses, block groups) using the mapping program put them all on the same scale. One caveat in using these measures relates to their inherent biases. The measures we used are more sensitive to populations of men who identify as gay or bisexual and hence to areas of the city where these men reside than they are to men who have sex with men but do not identify as gay or bisexual. This means we will include in our sample men who have sex with men but do not identify as gay or bisexual only if they live in areas of the city that have concentrations of gay and bisexual men. While we can't avoid missing portions of the target population (those who live outside the areas we selected), we most likely will be missing a larger proportion of men who do not identify as gay or bisexual but have sex with men than we will those who do identify as gay or bisexual.

As many in survey research are aware, working with unintegrated data sources like census data, zip codes, exchanges and telephone service boundaries can be quite time consuming, frustrating, and often futile. While it may be possible to acquire these sources in formats that are compatible, this course is often prohibitively expensive. Given the importance of this study in addressing issues in AIDS research among gay and bisexual men and with the resources we had available to us, we have been particularly attentive to devising sampling strategies that will capture the widest possible population of men residing in the selected cities. Only after the data are collected, will we have a better understanding of how well we succeeded.

References

- Binson, D., Michaels, S., Stall, R., Gagnon, J., Coates, T., and Catania, J., (1995), Prevalence and Social Distribution of Men Who Have Sex with Men: United States and Its Urban Centers," *Journal of Sex Research*, Vol. 32, No. 3, 245-254.
- Catania, J. A., Coates, T. J., Stall, R., Turner, H., Peterson, J., Hearst, N., Dolcini, M., Hudes, E., Gagnon, J., Wiley, J., & Groves, R. (1992a). Prevalence of AIDS-related risk factors and condom use in the United States. *Science*, 258, 1101-1106.
- Centers for Disease Control and Prevention. (1995) *HIV/AIDS Surveillance Report*; 7 (no. 2).
- Fay, R. E., Turner, C. F., Klassen, A. D., & Gagnon, J. H. (1989). Prevalence and patterns of same-gender sexual contact among men. *Science*, 243, 338-348.
- Holmberg, S. D. (1996) The Estimated Prevalence and Incidence of HIV in 96 Large US Metropolitan Areas. *American Journal of Public Health*, Vol. 5, (May), 642-654.
- Kanouse, D., Berry, S., Gorman, E., Yano, E., and Carson, S. (1991). *Response to the AIDS Epidemic: A Survey of Homosexual and Bisexual Men in Los Angeles County*. Santa Monica, CA: Rand.
- Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: The University of Chicago Press.
- Osmond, D., Page, K., Wiley, J., Garrett, K., Sheppard, H., Moss, A., Schragger, L., and Winkelstein, W. (1994) HIV Infection in Homosexual and Bisexual Men 18 to 29 Years of Age: The San Francisco Young Men's Health Study. *American Journal of Public Health*, December, 1994, Vol. 84, No. 12. 1933-1937.
- Rogers, S. M., & Turner, C. F. (1991). Male-male sexual contact in the U.S.A.: Findings from five sample surveys, 1970-1990. *The Journal of Sex Research*, 28, 491-519.
- Stall, R., Barrett, D., Bye, L., Catania, J., Frutchey, C., Henne, J., Lemp, G., & Paul, J. (1992). A comparison of younger and older gay men's HIV risk-taking behaviors: The Communication Technologies 1989 cross-sectional survey. *Journal of Acquired Immune Deficiency Syndromes*, 5, 682-687.
- Stall, R., Barrett, D., Bye, L., Catania, J., Frutchey, C., Henne, J., Lemp, G., and Paul, J. (1992) "A Comparison of Younger and Older Gay Men's HIV Risk-Taking Behaviors: The Communication Technologies 1989 Cross-Sectional Survey," *J. of Acquired Immune Deficiency Syndromes*, 5, No. 7:682-687.
- Waksberg, J. (1978) Sampling methods for random digit dialing. *Journal of the American Statistical Association*, Vol. 73, No. 361:40-46.
- Winkelstein, Jr., W., Lyman, D., Padian, N., Grant, R., Samuel, M., Wiley, J., Anderson, R., Lang, W., Riggs, J., and Levy, J. (1987) Sexual Practices and Risk of Infection by the Human Immunodeficiency Virus: The San Francisco Men's Health Study. *Journal of the American Medical Association*, Vol. 257, No. 3, 321-325.