Journal of Counseling Psychology
2009, Vol. 56, No. 1, 23–31

0022-0167/09/$12.00 DOI: 10.1037/a0014587

Sampling Lesbian, Gay, and Bisexual Populations

Ilan H. Meyer and Patrick A. Wilson
Columbia University

Sampling has been the single most influential component of conducting research with lesbian, gay, and bisexual (LGB) populations. Poor sampling designs can result in biased results that will mislead other researchers, policymakers, and practitioners. Investigators wishing to study LGB populations must therefore devote significant energy and resources to choosing a sampling approach and executing the sampling plan. The authors describe probability and nonprobability sampling methods used in LGB populations and critically discuss the advantages and disadvantages of the sampling methods they review. The authors conclude that no single sampling methodology is correct or incorrect for use in LGB populations; rather, researchers must evaluate advantages and disadvantages of each sampling methodology in the context of the specific research question and the research design.

Keywords: sampling, lesbian, gay, methodology

Sampling of study participants has probably been one of the most important methodological factors influencing the evolution of research on lesbians, gay men, and bisexual men and women (LGB). Results from studies on the health and psychological well-being of LGBs can be biased by the sampling methods used by the investigators. Perhaps the most notorious examples come from early 20th-century studies in which prison and clinical samples were used to paint LGBs as mentally and morally deficient, supporting the received view that homosexuality is pathological (Morin, 1977). Beginning in the mid-20th century, studies emerged that began to repudiate such views. Evelyn Hooker's studies of the 1950s (Hooker, 1957) showed that gay/bisexual men were as healthy as heterosexual men. Hooker's innovation, to a great extent, was the use of samples of generally healthy gay men. Such studies, along with political and social changes of the 1960s and 1970s, led to the declassification of homosexuality as a mental disorder in 1973 (Bayer, 1981).

In the 1970s, Bell and Weinberg (1978) used a variety of sources in LGB communities in the San Francisco Bay area in their innovative sampling approach to recruit study participants. Using this sample, the authors demonstrated that LGBs sampled from the community are different from the picture of the gay person that clinical samples had portrayed. But most studies of the 1970s and 1980s relied on opportunistic sampling, using limited sources of recruitment, thus increasing the potential for bias. For example, trying to estimate prevalence alcoholism in LGBs, Israelstam and Lambert (1986) reviewed 32 published studies. But of these, only 9 (28%) studies were based on community rather than clinical samples, and most of these used only one recruitment source in the gay community. Similarly, when the Centers for Disease Control (CDC; 1987) wanted to estimate prevalence of HIV, they relied on existing studies. But 34 (72%) of the 47 early HIV-prevalence studies the CDC had identified were based on clinical samples. In an effort to improve estimates of the emerging AIDS epidemic, large-scales studies were developed, and researchers began to use more sophisticated sampling methods, including venues in the gay community (Kaslow et al., 1987; Martin, 1987) and household probability sampling (Winkelstein et al., 1987). Recently, a new wave of studies advanced researchers' understanding of LGB populations by using randomly selected samples drawn from the general population to understand the subsamples of LGB individuals therein (e.g., Cochran, Sullivan, & Mays, 2003; Gilman et al., 2001; Sandfort, de Graaf, Bijl, & Schnabel, 2001).

In this article, we aimed to provide researchers, readers, journal editors, and reviewers with principles for thinking about sampling in LGB populations. We describe important sampling approaches used in LGB studies and discuss their strengths and weaknesses. We do not intend to provide a comprehensive review of sampling methods; for greater detail, the reader should refer to general and LGB-specific discussions (e.g., Binson, Blair, Huebner, & Woods, 2007; Rothblum, 2007).

Challenge in Sampling LGB Populations

A *sample* is a "subset from a larger population" (Sudman, 1976, p. 11). From this definition it is clear that before a researcher begins to design a sampling plan, he or she must have a good definition of the population. But here lies the first problem for researchers of LGB populations: The population's definition is elusive. First, all LGB individuals do not define themselves as LGB until some developmental tasks along the *coming out* process have been achieved (Eliason & Schope, 2007). This means that at any point some people who answer truthfully that they are not LGB will at a later point define themselves as LGB. Furthermore,

Ilan H. Meyer and Patrick A. Wilson, Mailman School of Public Health, Department of Sociomedical Services, Columbia University.

The present article was supported, in part, by National Institute of Mental Health Grant R01MH066058 awarded to Ilan H. Meyer. We thank Curtis Dolezal, Bonnie Moradi, Jonathan Mohr, and Tibor Palfai for valuable comments on an early version of this manuscript.

Correspondence concerning this article should be addressed to Ilan H. Meyer, Mailman School of Public Health, Department of Sociomedical Services, Columbia University, 722 West 168th Street, Room 533, New York, NY 10032. E-mail: im15@columbia.edu

because of cultural diversity, some people who engage in same-sex behavior, who may be considered by others as sexual minorities, and who may be of interest to the researcher, would not identify themselves as LGB, nor consider themselves a sexual minority by any name, regardless of the researcher's definition.

Even if we disregard these exceptions, definitions of sexual minorities vary: Several populations may be defined. Researchers have distinguished among sexual identity, sexual behavior, and attraction (Sell, 2007). Although these overlap—that is, a person who is attracted to same-sex individuals may also have sex with same-sex individuals—this overlap is not great, only among 15% of women and 24% of men do the three categories overlap (Laumann, Gagnon, Michael, & Michaels, 1994). Even within each of these categories, varied groups can be defined. Identity labels—and even whether a person uses an LGB identity label at all—vary across generations, racial/ethnic groups, geographical regions, education levels, and other group characteristics. Behavioral definitions, which rely on seemingly objective and clear criteria (often asked as "Have you had sexual relationships with men, women, or both men and women?"), also vary. For example, researchers have referred to different time periods for assessing sexuality (past year, past 5 years, since age 18, and ever). Because more people had same-sex sex in adolescence, defining sexual orientation as sexual behavior "ever" includes more people than defining it as "past year." This can lead to significantly different estimates: Laumann et al. (1994) found that 42% of all men who had ever had same-sex contact had none after age 18.

Finally, even if issues of definition were not a problem, sexual minority identity is highly stigmatized. Despite improvements in the social environment of LGB individuals, LGBs have much to lose from disclosing their sexual minority status. LGBs can be legally discriminated against in employment (e.g., "Don't Ask, Don't Tell" in the U.S. military) and are subject to rejection and violence (Herek, in press). With much to lose, LGBs may not be willing to disclose their identity to researchers.

Another problem for studies of LGB individuals is that even if researchers agreed on a population definition, they cannot find descriptive statistics about the characteristics of this population (e.g., its racial and educational demographics) because the LGB population has never been appropriately enumerated (Blair, 1999). The U.S. Census, which provides a description of the U.S. population and a benchmark for most population sampling, does not include information on sexual orientation. With no proper description of the LGB population, researchers cannot evaluate whether a sample is representative of the population—a great handicap for determining generalizability and assessing a study's results.

What all this suggests is that the most important part of any sampling design is conceptual: defining the population we want to study. Once the population is defined conceptually, a sampling frame has to be defined. It is the sampling frame, not the population, which is used for sampling. We can think of the population that the researcher defines as the conceptual population and the sampling frame as the operationalization of this construct. The sampling frame is the best approximation of but it is not identical to the conceptual population. Typically, some compromises must be made in order to arrive at a workable operationalization of the population. This compromise could lead to a redefinition of the conceptual population. For example, national population samples, whose conceptual construct is the entire adult U.S. population, typically exclude institutionalized individuals (in jails, mental institutions, and nursing homes) from the sampling frame, thus redefining the population of interest. Consider, for example, a researcher who aims to study satisfaction of LGB clients with a mental health clinic. At first glance, the population definition seems obvious, but it is not. Does the population of interest include all clients who ever contacted the clinic? Only those who have made an appointment in the past year? Only those who attend regularly? Each of such definitions has a different meaning and results in a different sampling frame. Of course, defining the population and forming a sampling frame of LGBs in the general population is more complicated than it would be in a clinic, where the population is relatively confined.

So what is the correct definition of the LGB population? The answer depends on the purpose of the study: A researcher who is interested in risks for HIV/AIDS among men who have sex with men (MSMs) might focus on behavioral definitions because behavior affects risk exposure regardless of personal identity. A researcher who is interested in developmental milestones of gay youth might focus on identity definitions because development of a gay identity is a core task facing the youth. Thus, there is not one answer to the question. It is the researcher's intellectual responsibility to answer this question with reasoned justification. The researcher must define the population of interest on the basis of the study's objectives and its underlying conceptual framework.

Below, we review probability and nonprobability sampling methods that are most often used to sample LGB populations. We provide a brief description of each method and discuss its strengths and limitations. With this, we hope to help researchers discuss explicitly the applicability of a chosen sampling methodology for the purpose and context of a research project. We begin our review with probability sampling methods, which are the most costly of the methods we describe, and then discuss nonprobability sampling methods that are more easily obtainable.

Sampling Methodologies

Probability Samples

Probability samples (sometimes referred to as *random samples*) are the gold standard of sampling for survey research. Most important, probability samples allow generalization of the results to the population from which the sample was drawn. This is necessary when researchers attempt to estimate population parameters, such as the prevalence of a disease in a population. *Probability sampling* means that every person in the population has a known nonzero probability of being included in the sample (Sudman, 1976). A zero probability of being included in a sample means that some individuals are hidden from the researcher and cannot be included in the sample.

There are different techniques of random sampling. In *simple random sampling*, an entire population is targeted with equal probabilities for selection of each person. Most often, though, more complex techniques are used, such as *stratified* and *cluster sampling*, that differentially target segments of the population. In such designs, the probability of being included in the sample varies for different subgroups of the population (e.g., women), but because the researcher controls the inclusion probabilities, they are known. When the probability of being included is known, even when dissimilar probabilities are used for different segments of the population, the researcher can adjust the results so that correct total population estimates can be reconstituted.

The greatest disadvantage of probability sampling for LGB population is its cost. Because LGB-identified individuals are a minority in the general population (present estimates hover around 1%–4%), collecting a probability sample of LGBs across the U.S. population would be prohibitively expensive. Therefore, many probability samples of LGB individuals are subsets of large-scale studies in the general population (see review in Meyer, 2003). These studies have provided important descriptions of the LGB population as a whole, for example, regarding prevalence of mental disorders among LGBs. But such studies include too few LGBs to address questions about variability within the LGB community, such as variation across racial/ethnic groups or socioeconomic status (Meyer, Dietrich, & Schwartz, 2008).

To obtain larger samples of LGBs while reducing the cost of probability sampling, researchers have targeted geographic areas with greater density of LGBs ("gay neighborhoods"). For example, the Urban Men's Health Study of gay/bisexual men (Catania et al., 2001) targeted specific neighborhoods in four large metropolitan areas. Within the selected areas, the researchers used random-digit dialing (RDD), a telephone sampling technique. Even when targeting these high-density neighborhoods, the researchers had to attempt 53,050 phone numbers to reach the recruitment goal of 915 interviews in one of the study sites (Binson et al., 2007). Although sampling in targeted gay neighborhoods is an effective way to circumvent the challenges of locating LGB persons in the general population and, thus, reducing the cost of probability sampling, it is not an ideal approach to sampling. Targeted neighborhood sampling makes sense only when the population of interest is defined as individuals residing in these neighborhoods. Typically, though, researchers want to understand LGB people more broadly. When sampling LGB neighborhoods, researchers explicitly or implicitly assume that individuals in their sample are similar to LGBs not residing in such neighborhoods. But this is not a safe assumption: Compared with nonresidents, LGBs residing in LGB neighborhoods probably are more strongly identified as LGB, have stronger ties with and social supports within the LGB community, are of higher income and social class (because such neighborhoods tend to be more expensive neighborhoods), tend to be immigrants to the city rather than natives, and the like. Therefore, although the greatest benefit of probability sampling is in generalization, the modification to targeted neighborhood compromises this main benefit.

It is important to note that probability sampling does not remove all potential sampling biases. A great source of bias comes from differential response rate among subgroups of the population. In general, participants in survey research tend to be of higher education and income than nonparticipants. Additionally, some individuals may not be reached at all because of the sampling methodology even when they are part of the defined population. For example, in door-to-door household sampling, there is greater likelihood to miss individuals who work outside the home than those who stay at home. To the extent that individuals who are underrepresented in the sample are different from those who are included in the sample, inference errors would ensue. To demonstrate this problem with a hypothetical example, consider the following: A researcher may reasonably argue that an RDD sample, compared with a sample recruited from the LGB community in the city, is worth the added cost because it can capture LGB individuals who are less affiliated with the community and therefore less likely to be recruited in community venues. Suppose that when calling potential respondents, the research team identified the study as a study about LGBs and sought to motivate potential respondents by saying that by volunteering, participants would help the LGB community. It is possible that this framing of the study would motivate people with greater connection and affiliation with the LGB community, leading to greater response among them compared with less affiliated individuals. Inadvertently, due to response bias, the researcher may have oversampled affiliated LGB individuals—the exact pattern he or she had attempted to avert by using RDD.

Nonprobability Samples

For most investigators, the cost of LGB probability samples is prohibitive, requiring development of alternative sampling techniques. *Nonprobability* sampling refers to any sampling technique for which the probability of a person being selected into the sample is unknown. This means that in nonprobability sampling, some people of the desired population may not be included in the sample, and other people may be overrepresented. But unlike stratified or cluster probability sampling, in which a researcher determines the probabilities of inclusion in the sample, in nonprobability sampling the extent of over- or underrepresentation is unknown. Therefore, there is no way of adjusting for noninclusion, potentially biasing results. Just as with planning a probability sample, the researchers who use nonprobability sampling must explicitly define the population of interest and carefully match a sampling procedure to this definition. Because nonprobability sampling techniques can be less regimented than probability sampling techniques—the latter follow clear randomization procedures—and because there are many potential pitfalls in nonprobability sampling, very careful planning must precede nonprobability sampling. One of the most important tasks facing the researcher is to anticipate possible sampling biases and design methods that can minimize such biases.

Probability samples are always necessary when researchers aim to estimate population parameters such as prevalence of disease, attitudes, and voting patterns. Consider a researcher who wants to assess LGB clients' attitudes toward a mental health clinic's. Without a probability sample, the researcher will not be able to get an unbiased estimate of the proportion of the clinic's LGB clients who have a favorable view of the clinic. Assume that the clinic has Black and White LGB clients and that the Black clients had a less favorable view of the clinic than did the White LGB clients. If the researcher included in her or his sample a larger proportion of Black clients than their representation in the clinic, then she or he would get a biased estimate of the overall clinic's LGB clients' attitudes. But nonprobability samples are a good alternative when estimating population prevalence is not a research focus. For example, studies that aim to generate hypotheses may be less concerned with whether the magnitude of associations correctly reflects the associations found in the population and more concerned with whether and how certain variables relate to one another. This is often the case in psychological research, in which researchers study theoretically driven hypotheses. The researcher in our example could test hypotheses about clients' characteristics that are associated with a favorable view of the clinic. For example, she or he may ask whether the clinic is as welcoming to Black as it is to White LGB people and test the hypothesis that Black racial/ethnic identity is associated with less favorable attitudes about the clinic than is White identity. Using a nonprobability sample, the researcher would get a good idea about differences between Black and White clients' attitude and would be able to conclude correctly that Black clients have worse attitudes toward the clinic than do White clients, even if she or he cannot correctly estimate the proportion of Black and White clients in the clinic.

However, it is important to note that nonprobability sampling bias can affect even hypothesis testing. That could happen if a sample is so uncharacteristic of the population that even relationships among variables are inaccurately represented. For example, suppose a clinic researcher wanted to test the hypothesis that higher internalized homophobia may be related to projected negative attitudes toward the clinic. Suppose the researcher collects her sample one Tuesday afternoon and recruits a sufficient number of study participants. Coincidentally, unbeknownst to her, that Tuesday was the day that the clinic hosted its support group for people with coming out difficulties. The sample may be biased because of the unique nature of this group: Clients attending may have higher levels of internalized homophobia than the average clinic client. But why is this a problem? We already stated that the researcher is not attempting to estimate population parameters—she is not using this sample to estimate the level of internalized homophobia among the clinic's clients. Still, because the group clients all have high levels of internalized homophobia, this variable may be invariant. With insufficient variability, the researcher will find no relationship between internalized homophobia and attitudes toward the clinic and falsely reject the hypothesis. But because the sample's invariability does not represent the true variability in the clinic population, the conclusion is wrong. The findings do not represent the relationships the researcher would have found in a better (unbiased) sample of the clinic's clients.

Another pitfall of nonprobability samples involves volunteer bias. This bias relates to the special characteristics of respondents who volunteer to participate in a study. Such bias may be related to the special interest that volunteers have in the topic under study. For example, studies of eating disorders in gay/bisexual men have tested the hypothesis that gay/bisexual men have greater body dissatisfaction and, therefore, higher prevalence of eating disorders than do heterosexual men. Many studies recruited samples by advertising that they sought volunteers to study eating disorders. It is likely that such advertising attracted people with great interest in the topic, an interest motivated by personal struggles around eating. Such a sample of volunteers probably represents quite a unique experience around eating. To the extent that the sample is different from the general population of gay/bisexual men to which researchers generalized their results, the study results could introduce a sampling bias. Such sampling bias could affect not only estimates about prevalence of eating disorders but also inferences about other hypotheses. For example, in the highly selective sample of gay men with concerns about disordered eating, they may have little variability in body image, and, therefore, no association would be found between body image and disordered eating, even if in the total population this relationship were true. Such volunteer effect could cause bias in probability samples as well, as demonstrated above with the case of RDD respondents who were highly affiliated with the LGB community. But it is a more serious threat in nonprobability samples, in which the researcher has less control over the sampling procedure.

Both of these examples emphasize that thinking about potential pitfalls in the sampling procedure is an important task facing the researcher who designs a nonprobability sampling procedure. Through careful planning and foresight about potential biases, researchers can find solutions and minimize bias. In our first example, better familiarity of the clinic could have prevented sampling from the coming out support group. In the eating disorders example, careful wording of literature about the study could have prevented or reduced volunteer effects. Therefore, a good familiarity with the community—including the use of ethnography, theoretical consideration of potential sources of bias, and sound strategies to minimize potential bias—are all important elements of nonprobability sampling methodologies.

Nonprobability samples are sometimes referred to as *convenience* samples. We prefer to reserve that term to truly convenient samples—samples that the researcher uses because they happen to be readily available, such as the drop-in counseling group in the example above. Other examples include students in a classroom, a clinician's own clients, or attendants of a local bar. Most often, such convenience sampling is done without sufficient consideration of the conceptual definition of the population and with no careful consideration of potential biases. But for most nonprobability sampling procedures, "convenience" is a misnomer; nonprobability sampling requires very careful consideration, design, and execution of the sampling plan. There are many nonprobability sampling techniques, each with its own strengths and weaknesses. We discuss four methods used in LGB studies: sampling in LGB community venues, time-space sampling, respondent-driven sampling, and Web-based sampling.

Sampling in LGB community venues. One of the most frequently used methods for recruiting LGB individuals is through sources in the LGB community. Sampling in the community has a long history in modern LGB affirmative research, as researchers used contacts in the community to access the population that was otherwise impossible to locate (Rothblum, 2007). *Community venues sampling* refers to methods ranging from samples of convenience, such as sampling in a community organization, to more sophisticated methods described below. One of the strongest critiques of community venues sampling is that researchers using this approach can only reach LGB persons who partake in the LGB community, overlooking individuals who are not identified as LGB. Moreover, as this critique goes, individuals who do not partake in the LGB community are different from those who do; therefore, results from samples recruited in the LGB community are biased. For example, researhcers have observed that LGBs with high levels of involvement in the gay community have different psychological and risk profiles than those not involved (Ramirez-Valles, 2002). If the researcher's aim is to understand risk among MSMs, including those who are not gay identified, sampling in gay venues will lead to biased findings.

This critique is valid, but it is often taken to mean that community venues are inappropriate as sampling sources and should be avoided. That conclusion, we think, is mistaken: The appropriateness of a sample can only be judged in the context of the research hypotheses and the related definition of the population. A sample recruited from the LGB community would be relevant to a host of research questions that concern the lives of LGB persons who make up this community, including questions about health and well-being, relationships and intimacy, social support, children and parenting, identity development and identity conflict, and political affiliations and attitudes. Of course, researchers who use such samples need to generalize their results appropriately to the population of affiliated LGB individuals.

A related critique of community-based sampling is that individuals' (unknown) probability of inclusion in the sample is proportional to their level of activity in the LGB community—the more often one participates in a community group, the more likely one is to be sampled if community groups are a source of sampling. To address this, researchers can use a secondary sampling procedure. In *snowball sampling*, respondents selected from community venues are asked to nominate potential participants from among their social network. These nominees are invited to participate in the study; in turn, they too are asked to nominate individuals from their own social networks. By reaching into social networks, researchers presumably sample individuals who are less likely to be contacted through the primary outreach in community venues. It should be noted, however, that including members of social networks imports new biases of its own: Members of social networks are likely to be more socially connected than individuals not connected to social networks, and social connectedness can correlate with numerous variables of interest (Harry, 1986).

Another problem with venue sampling is that bias can be introduced from a venue used to recruit study participants. We discussed studies in which a single source for sampling was used, such as gay bars, and the potential bias imported into these studies because of the special characteristics of the venue. All venues attract more of some type of person than others. If the characteristics that are associated with a venue are correlated with the variables of interest in the study, then the results of the study may be biased. Bias can be introduced not only by using select venues but also by sampling too many individuals from any one venue. For example, in an effort to enroll young Black MSMs into a survey, Gay Men's Health Crisis (GMHC) offered free admission to a New York City club and recruited 1,500 participants at the event (Altman, 1999). To the extent that the 1,500 participants share characteristics—and by virtue of their attendance at this club, it is likely that they probably share characteristics such as age, social class, familiarity with GMHC, and the like—their large number would overwhelm responses from any other type of participants in the study. Similarly to our example above about internalized homophobia in the sample recruited from the mental health clinic's drop-in group, in this case, generalization to Black MSMs would be misleading because the study sample seems to represent a very narrow subgroup of the population of interest.

To prevent such bias, researchers have attempted to balance representation of participants from various sources of recruitment and avoid altogether venues that could introduce serious bias. For example, Meyer, Schwartz, and Frost (2008) used ethnographic methods to prepare a sampling frame of New York City's LGB communities. The venues were classified by type (business, public spaces, LGB groups and organizations, etc.), population (Latina lesbians, Black men, etc.), location (the researchers included venues in four of New York City's boroughs), and other characteristics. To reduce bias, the researchers excluded any venue that was not open to the general LGB public or subgroups therein. Researchers excluded venues where attendees were likely to have characteristics that correlate with the main variables of interest in the study—mental health and stressful events (e.g., treatment venues such as 12-step programs and venues that catered to people who experienced a special life event such as antigay violence). From the sampling frame, the researchers purposefully selected venues for recruitment, aiming to maximize variability of the types of venues. Quota sampling was used to ensure that the study participants include sufficient representation of men and women, racial/ethnic variability, and the desired age distribution as predetermined by power analysis. To avoid overwhelming the sample with respondents recruited at any one venue, the researchers limited recruitment by venue type/time. Although such methods can reduce potential bias, they provide no method of quantifying and statistically correcting for sampling bias, such as when using probability sampling methods.

Time-space sampling. Time-space sampling (TSS) is a nonprobability sampling approach that attempts to assert greater control on potential sampling biases. In particular, TSS helps to prevent biases related to the convenience of some sites over others and biases related to oversampling from one site compared with another site because of serendipitous circumstances or misguided efforts to improve sample size (e.g., the GMHC study's recruitment of 1,500 young Black MSMs in a gay nightclub). The first step in TSS involves identifying and selecting venues in which the target population can be found. With the help of ethnography, researchers select venues that represent the desired population of venues (preferably all relevant venues). That is, if MSM sexual risk is being studied, then as many venues where MSMs can be contacted as possible should be included. This is because different types of MSMs are likely to be found in different venues. This step is similar to LGB community venues sampling, but unlike the venue-based sampling, in TSS the researchers also record the patterns of attendance at the venues. To do that, research assistants visit the venues at various times of the day and collect information about the number of eligible persons (e.g., MSMs) who visit the venue at different times of the day and night. From these data, researchers create the sampling frame of time-space units. A *time-space unit* is a day of the week and time of the day when participants visit a particular venue (e.g., 30 people were at Joe's bar on Tuesday between 6 p.m. and 10 p.m. etc.). This is important because sampling bias can be introduced not only by using select venues but also by sampling too many individuals from a venue at any one time. From this sampling frame, researchers randomly select time-space units for recruitment of study participants. The number of participants recruited at a time-space unit is proportional to the frequency of attendance observed at the venue at the specific time. Some of the difficulties in TSS include recruiting respondents at sites where they are engaged in recreational or other activities and therefore not as willing to participate in a study.

Also, it is challenging to identify eligible respondents both when creating the sampling frame and when recruiting participants. This is especially difficult at sites that do not cater to the target population exclusively (e.g., a public park that attracts both MSM and non-MSM individuals). Another difficulty is counting or recruiting a person only once when people leave and reenter the venue. These and other challenges can make TSS difficult to execute (Stueve, O'Donelle, Duran, San Doval, & Blome, 2001).

Thus, TSS can be a robust sampling approach when selection of venues and balance among venues is of primary importance. But TSS can be complex to execute and requires resources that may be outside the reach of many researchers. In some areas, where there are not many venues where LGB people congregate, or where the LGB people cannot be readily identified, it makes little sense to attempt TSS. A distinction of TSS versus community-based sampling is that TSS samples a population defined by the sites of recruitment (Binson et al., 2007). That is, in venues sampling, the researcher defines the population and attempts to find its members in venues, hoping that members of the population who are found in the venues are not grossly different than those not in the venues. In TSS, the population is defined more strictly as individuals who attend the venues. For HIV/AIDS studies, in which TSS has been used, this has meant individuals who through their participation in venues (e.g., bars, bathhouses, cruising parks) engage in HIV-related risk. But sampling venues rather than populations are not always appropriate. Even when appropriate, bias can be incorporated into TSS if the researchers did not compose a sampling frame that accurately represents the population of venues where the population of interest congregates, or if the population of interest, or a portion of it, is not accessible at venues. As Xia and colleagues (2006) showed, gay-identified venues attract different subgroups of MSMs than nongay venues, and men's sexual risk profiles differ by type of venue.

Respondent-driven sampling. Respondent-driven sampling (RDS) was designed for sampling hidden populations that are difficult or impossible to sample using probability sampling methods (Heckathorn, 1997). Hidden populations are populations about which important characteristics are unknown and whose members would not be identified readily because of stigma. An advantage of RDS is that it does not rely on venues as do TSS and community-based sampling; instead, RDS relies on social networks. Both RDS and snowball sampling assume that members of the population are best able to reach their peers through social networks, but unlike snowball sampling, a more exacting procedure is used in RDS: A system of incentives motivates network members to participate in the study in a way that a field outreach worker, using the less structured snowball sampling method, cannot. A central concept of RDS is *group-mediated social control*, which suggests that compliance with behaviors like participation in a study can be motivated by peer pressure. The most important purported benefit of RDS, and its advantage over other community-based sampling methods, is that researchers using RDS can make unbiased population estimates similar to probability sampling methods (Salganik & Heckathorn, 2004). As we discuss below, however, this claim has been contested by critics.

A two-tiered incentive system (typically a monetary reward) is used in RDS. The first step in RDS is the identification of an initial wave of respondents—called *seeds*—who fit eligibility criteria for the study. The seeds are then given recruitment *coupons* that they would use to invite peers who meet inclusion criteria to participate in the study. When a person is enrolled in the study through referral from a seed, he or she presents the coupon that identifies the seed. The seed then gets a reward for the successful referral (secondary incentive). In turn, persons who are enrolled as study participants are offered the same dual incentive plan as the seeds—they are rewarded for their participation in the study as well as for bringing new participants from their own social networks into the study (they are called *recruiters*, to differentiate them from the seeds). The number of coupons provided to the seeds and recruiters is predetermined and limited to prevent oversampling by seeds and recruiters with larger personal networks than others. This process continues until the desired sample size and composition have been achieved.

RDS can be difficult to implement. One difficulty is that because of the strong reliance on incentives, some individuals—especially in impoverished populations—may be motivated to participate in a study fraudulently even if they do not meet eligibility criteria. To guard against such violations, Heckathorn (1997) suggested that the respondent's traits that qualify him or her to participate in the study be objectively verified. RDS was originally developed for use with injection drug users (IDU), for whom verification is somewhat easier (e.g., by observing track marks on injection sites). In studies of LGBs, objectively verifying an LGB identity or behavior is impossible, but it is important to develop some method of scrutiny of enrollment eligibility (e.g., by asking how recruits know the seed). Another problem is participants attempting to complete the survey (and gain reward) more than once.

The key purported advantage of RDS is that researchers who use this method can make unbiased population estimates (Salganik & Heckathorn, 2004). This is a key advantage of probability sampling; if RDS could provide unbiased population estimates without the need for a sampling frame, then this would indeed be of great benefit for studies of LGB populations. Probability sampling allows estimation of population parameters because a probability sample is an unbiased representation of the population. In RDS, estimation is indirect: The sample allows estimation of networks properties, and estimation of networks properties allows estimation of the population from which the networks were drawn (Salganik & Heckathorn, 2004). But researchers have critiqued the claim that RDS can arrive at unbiased population estimates, saying that the assumptions underlying RDS, which make probability estimates feasible, are rarely true in the case of LGBs (Binson et al., 2007). A core assumption of RDS is that "the network of the hidden population forms one connected component" (Salganik & Heckathorn, 2004, p. 210). Salganik and Heckathorn (2004) claimed that this is a reasonable assumption for "many hidden populations" (p. 210), but we disagree with regard to LGBs. The authors give as examples of networks that form a connected component jazz musicians and IDUs. This is in contrast with tax evaders whose networks do not form a connected component. We suspect that the connected components assumption cannot be easily accepted for LGBs. Unlike jazz musicians and IDUs, the diversity of LGBs makes it hard to conceive that the networks of LGBs form a connected component. Even within one city, there would be very little contact between networks of what we would want to capture under *LGB populations*. We believe that in this sense, LGBs are more similar to tax evaders, who do not see each other as peers, than they are similar to jazz musicians, who, through their shared avocation, often do move in similar circles that form a connected component.

Heimer (2005), critiquing Ramirez-Valles, Heckathorn, Vazquez, Diaz, and Campbell's (2005) RDS sample of Latino gay men, noted that several other assumptions, related to the representativeness of network sizes among seeds and the potential for homophily (i.e., preferential, or nonrandom, recruitment of network members) among seeds and recruiters, were likely not achieved by the researchers. Therefore, Heimer concluded that it is inaccurate to describe the sample as representative of Latino gay men. Furthermore, critics claim these and other assumptions are often violated in research using RDS in LGBs, leading to gross misrepresentations of findings (Binson et al., 2007).

Web-based sampling. Web-based survey methods include both sampling of study respondents on the Web and the use of the Web for delivering questionnaires to study participants regardless of how participants were sampled. Our focus here was on the use of the Web to recruit samples of LGB individuals. Using the Web to sample LGBs for research has great benefits and the potential to address gaps in present sampling methodologies. Most significantly, using the Web, researchers can reach populations that have been overlooked in LGB research, such as LGBs in rural areas, in towns and villages where only small numbers of LGBs reside, and even internationally. Web-based samples, like samples from other community venues, can be biased by special characteristics of the venues in which participants are recruited. Some of the procedures described previously, like quota sampling, TSS, and respondent-driven sampling, can be used in conjunction with Web-based sampling. To improve the selection of potential Web sites as recruitment venues, it is important that researchers conduct *cyber-ethnography* to characterize the Web environment and the different audiences who use various Web sites (Carballo-Dieguez et al., 2006).

As we show below, Web-based sampling also has many pitfalls. As with any other methodology, Web-based sampling should be used only after consideration of its strengths and limitations for a particular research project. Among the limitations of Web-based sampling is that, despite high use of the Internet in the United States, about 27% of Americans do not use the Internet (Madden, 2006). Most importantly for sampling, there are significant differences between LGB with and without Internet access. In the general population, a *digital divide* exists: Americans with Internet access are younger, have a higher socioeconomic status, and are less likely to be racial/ethnic minorities than those without Internet access (Fox, 2005; Gosling, Vazire, Srivastava, & John, 2004), and gay/bisexual men Internet users engage in higher HIV-related risk behavior than nonusers (Liau, Millett, & Marks, 2006). Thus, sampling LGBs on the Web could exclude distinct segments of the LGB population and therefore yield an unrepresentative sample.

Recruitment of study respondents on the Web may be *active* or *passive*. In active recruitment, researchers engage with Internet users and ask them to participate. This approach has the advantage for researchers of establishing personal contacts with potential participants and allows participants to ask questions about the study. However, many Web sites, particularly those that are sexually explicit or deal with provocative subject matters, prohibit online solicitation of study respondents. Passive recruitment uses advertisement such as Web site banners (Pequegnat et al., 2007). Passive Web-based sampling may be the most efficient way to reach large numbers of LGBs, though it provides the lowest response rate. For example, Rosser and colleagues (2008) used an ad placed on the popular Web site gay.com to recruit Latino gay/bisexual men/MSMs. Over 3 weeks, there were over 47 million views of the ad and roughly 33,000 (0.07%) people followed the ad. Despite this extraordinarily low proportion of responders, the researchers obtained over 1,000 completed questionnaires.

Another Web-sampling approach involves obtaining a sample from already collected panels of Internet users. Different research panels are maintained by large survey/market research firms such as Harris Interactive and Ipsos Research. Of course, these are panels of volunteers not randomly selected to represent the U.S. or LGB populations. The response rate from some of these panels can be quite low (less than 10%), further exacerbating the question about representativeness of the sample and adding to concern about sampling bias. One of the most sophisticated Web-based samples of LGBs is a panel developed by the firm Knowledge Networks. Knowledge Networks developed a probability-based panel designed to represent the U.S. population. This panel was not recruited on the Web. We include it here because it provides the benefit of reaching distant and elusive populations using the Internet while overcoming the limitation of nonrepresentativeness of Internet users. Knowledge Networks recruited potential study participants using RDD methods from households representing the U.S. population. If a person were selected but did not already have Internet access, Knowledge Networks provided him or her with Internet access, bridging the digital divide. Herek (in press) used a Knowledge Networks sample of 662 LGBs to study the prevalence of exposure to hate crimes and stigma-related experiences.

In summary, the Web can be a great resource for reaching dispersed and otherwise hard-to-reach populations. Using the Internet, researchers have obtained samples that are difficult to get using other methods: Rosser, Oakes, Bockting, and Miner (2007) studied what may be the largest sample of transgender persons in the United States to date—over 1,200 individuals; Wilson et al. (in press) studied MSMs who engage in bareback sex; and Bowen (2005) studied rural MSMs. But a great concern in Web-based samples is generalizability and selection bias. Of course, as we have seen above, these limitations are not unique to Web-based sampling. But researchers who use Web-based sampling typically do not know how many people viewed their solicitation, what motivated participants to respond, and how different or similar is the sample to the conceptual population of interest.

Conclusions

Several characteristics of LGB populations make it challenging for sampling. Chief among these are that the population is difficult to define conceptually, that LGBs are stigmatized and may resist disclosure of their sexual behavior or identity to researchers, and that people may apply a variety of identity labels or no identity labels at all to themselves. More significantly, factors that cause variation in sexual orientation expression, definition, and identification are not randomly distributed. Sexual expression, identification, and disclosure may be associated with socioeconomic, cultural, social, and personality characteristics that, in turn, correlate with outcome variables such as well-being and distress. Therefore, sampling bias could have a great effect on inference from research results.

These challenges do not mean that good LGB research is impossible to do with moderate resources. Many of the problems we discussed are problems that can be addressed with careful consideration. Researchers must carefully weigh the strengths and weaknesses of sampling approaches and assess what approach best matches the research questions. Of course, this is always true in sampling design (Sudman, 1976), but it is especially important in the case of LGB populations, in which various population definitions are plausible. In choosing one sampling approach over another, the researcher must be satisfied that the advantages of the chosen approach outweigh its disadvantages for the specific research purpose.

Researchers must define the population before they can move on to determine what sampling technique to use (Sell, 2007). Too often this is not done explicitly; instead, researchers proceed with implicit definitions of the population. Worse, researchers sometimes define the population retrospectively, based on the sample they had obtained. Such a researcher might conclude that the results apply to the types of people who were included in the sample (e.g., university students) even though the research questions were not specific to this population and the conclusions imply more general applicability. Such practice is misleading. Journal editors, reviewers, and readers must critically examine the research conclusions of a study in view of these limitations. Readers and researchers should not be influenced by the *appearance* of methodological complexity. A sample should be assessed only in the context of the research questions and inferences it makes, not by any objective measure of sampling sophistication. Even a probability sample expertly conducted may be the wrong sample if it is limited to a particular neighborhood or organization that does not represent the population about which the researcher infers. Such a sample is representative of the sampling frame, but because the sampling frame is not representative of the population of interest, the sample lacks external validity.

Researchers should beware confounding sampling and other methodological issues—researchers sometimes confuse sample size with sample representativeness (external validity). It should be clear that sample size affects power and statistical conclusion validity, not external validity. For example, a sample of 100 clients who attend a coming out support group session is no more representative of an entire clinic population than would be a sample of 25 clients who attended that session. Clearly, the problem with both is the opportunistic sampling; whether 25 or 100 respondents should be studied depends on the planned statistical analyses.

Somewhat similar is an opposite error: Sometimes researchers aim for a wholly representative sample—one that would represent the multitudes of subpopulations that make up LGB diversity. Such a researcher includes all possible subgroups (transgender people, older and younger people, all racial/ethnic groups, etc.). Although inclusion of diverse LGB populations should be an important goal for researchers, it can be misguided if the number of respondents in each subgroup is insufficient for subgroups-specific analysis. Power calculation should determine the minimum subgroup size required in a study so that meaning inference can be made about the subgroup. Including individuals who represent diversity in insufficient numbers would result in conducting analyses on the entire group regardless of subgroup affiliation, effectively obscuring diversity, or discarding data of individuals in small subgroups altogether. Researchers should focus their research on groups and subgroups that make sense scientifically and should sample a sufficient number of respondents in each subgroup to answer these questions. If and when diversity is not fully included—as would be the case in most studies—this deficiency should be highlighted and relevant questions discussed so that other researchers may fill the gap.

Challenges to conducting quality research on LGB populations could discourage researchers. But researchers should not shy away from studying LGB populations because of these challenges. Although we described some unique challenges, challenge itself is not unique to LGB research. The challenges of finding a good representative sample of LGBs are not fundamentally different from the challenge of finding good representative samples of other relatively rare groups (e.g., Jewish individuals, Latina older women, and Black youth). Researchers, reviewers, and journal editors should keep a critical eye when evaluating sampling methodology in LGB research. At the same time, they should not adhere to such strict guidelines that would thwart progress and impede gaining important knowledge about the lives of LGB people.

References

Altman, L. K. (1999, June 28). New York study finds gay men using safer sex. Retrieved November 17, 2008, from http://nytimes.com

Bayer, R. (1981). *Homosexuality and American psychiatry*. New York: Basic Books.

Bell, A. P., & Weinberg, M. S. (1978). *Homosexualities.* New York: Simon & Schuster.

Binson, D., Blair, J., Huebner, D. M., & Woods, W. J. (2007). Sampling in surveys of lesbian, gay and bisexual people. In I. H. Meyer & M. E. Northridge (Eds.), *The health of sexual minorities: Public health perspectives on lesbian, gay, bisexual and transgender populations* (pp. 375–418). New York: Springer.

Blair, J. (1999). A probability sample of gay urban males: The tale of two-phase adaptive sampling. *Journal of Sex Research, 36,* 39–44.

Bowen, A. (2005). Internet sexuality research with rural men who have sex with men: Can we recruit and retain them? *Journal of Sex Research, 42,* 317–323.

Carballo-Dieguez, A., Dowsett, G. W., Ventuneac, A., Reimen, R. H., Balan, I., Dolezal, C., et al. (2006). Cybercartography of popular internet sites used by New York City men who have sex with men interested in bareback sex. *AIDS Education and Prevention, 18,* 475–489.

Catania, J., Osmond, D., Stall, R. D., Pollack, L., Paul, J. P., Blower, S., et al. (2001). The continuing HIV epidemic among men who have sex with men. *American Journal of Public Health, 91,* 907–914.

Centers for Disease Control. (1987). HIIV-infection prevalence. *Mortality and Morbidity Weekly Report, 36,* 22–23.

Cochran, S. D., Sullivan, J. G., & Mays, V. M. (2003). Prevalence of mental disorders, psychological distress, and mental services use among lesbian, gay, and bisexual adults in the United States. *Journal of Consulting and Clinical Psychology, 71,* 53–61.

Eliason, M. J., & Schope, R. (2007). Shifting sands or solid foundation? Lesbian, gay, bisexual, and transgender identity formation. In I. H. Meyer & M. E. Northridge (Eds.), *The health of sexual minorities: Public health perspectives on lesbian, gay, bisexual and transgender populations* (pp. 3–26). New York: Springer.

Fox, S. (2005). *Digital divisions: There are clear differences among those with broadband connections, dial-up connections, and no connections at all to the internet*. Washingston, DC: Pew.

Gilman, S. E., Cochran, S. D., Mays, V. M., Hughes, M., Ostrow, D., & Kessler, R. C. (2001). Risks of psychiatric disorders among individuals reporting same-sex sexual partners in the National Comorbidity Survey. *American Journal of Public Health, 91,* 933–939.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59,* 93–104.

Harry, J. (1986). Sampling gay men. *The Journal of Sex Research, 22,* 21–34.

Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44,* 174–199.

Heimer, R. (2005). Critical issues and further questions about respondent-driven sampling: Comment on Ramirez-Valles, et al. (2005). *AIDS and Behavior, 9,* 403–408.

Herek, G. (in press). Hate crimes and stigma-related experiences among sexual minority adults in the United States: Prevalence estimates from a National Probability Sample. *Journal of Interpersonal Violence*.

Hooker, E. (1957). The adjustment of the male overt homosexual. *Journal of Projective Techniques, 21,* 18–31.

Israelstam, S. L. S. (1986). Homosexuality and alcohol: Observation and research after the psychoanalytic era. *Substance Use and Misuse, 21,* 509–537.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., & Rinaldo, C. R., Jr. (1987). The Multicenter AIDS Cohort Study: Rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology, 126,* 310–318.

Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. Chicago: University of Chicago Press.

Liau, A., Millett, G., & Marks, G. (2006). Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. *Sexually Transmitted Diseases, 33,* 576–584.

Madden, M. (2006). *Internet penetration and impact. Pew Internet & American Life Project reports*. Retrieved October 18, 2008, from http://www.pewinternet.org/pdfs/PIP_Internet_Impact.pdf

Martin, J. L. (1987). The impact of AIDS on gay male sexual behavior pattern in New York City. *American Journal of Public Health, 77,* 578–581.

Meyer, I. H. (2003). Prejudice as stress: Conceptual and measurement problems. *American Journal of Public Health, 93,* 262–265.

Meyer, I. H., Dietrich, J., & Schwartz, S. (2008). Lifetime prevalence of mental disorders and suicide attempts in diverse lesbian, gay, and bisexual populations. *American Journal of Public Health, 98,* 1004–1006.

Meyer, I. H., Schwartz, S., & Frost, D. M. (2008). Social patterning of stress and coping: Does disadvantaged status confer excess exposure and fewer coping resources? *Social Science & Medicine, 67,* 368–379.

Morin, S. F. (1977). Heterosexual bias in psychological research on lesbianism and male homosexuality. *American Psychologist, 32,* 629–637.

Pequegnat, W., Rosser, B. R. S., Bowen, A. M., Bull, S. S., DiClemente, R. J., Bockting, W. O., et al. (2007). Conducting Internet-based HIV/STD prevention survey research: Considerations in design and evaluation. *AIDS and Behavior, 11,* 505–521.

Ramirez-Valles, J. (2002). The protective effects of community involvement for HIV risk behavior: A conceptual framework. *Health Education Research, 17,* 389–403.

Ramirez-Valles, J., Heckathorn, D. D., Vazquez, R., Diaz, R. M., & Campbell, R. (2005). From networks to populations: The development and application of respondent-driven sampling among IDUs and Latino gay men. *AIDS Behavior, 9,* 387–402.

Rosser, B. R. S., Miner, M. H., Bockting, W. O., Ross, M. W., Konstan, J., Gurak, L., et al. (2008). *HIV risk and the Internet: Results of the Men's INTernet Sex (MINTS) Study*. Retrieved November 17, 2008, from http://www.springerlink.com.monstera.cc.columbia.edu:2048/content/ft42733730q32n70/fulltext.html

Rosser, B. R. S., Oakes, J. M., Bockting, W. O., & Miner, M. (2007). Capturing the social demographics of hidden sexual minorities: An Internet study of the transgender population in the United States. *Sexuality Research and Social Policy, 4,* 50–64.

Rothblum, E. (2007). From science fiction to computer-generated technology: Sampling lesbian, gay, and bisexual individuals. In I. H. Meyer & M. E. Northridge (Eds.), The health of sexual minorities: Public health perspectives on lesbian, gay, bisexual and transgender populations (pp. 441–454). New York: Springer.

Salganik, M. J., & Heckathorn, D. D. (2004). Sampling and estimation in hidden populations using respondent-driven sampling. *Social Methodology, 34,* 193–240.

Sandfort, T. G. M., de Graaf, R., Bijl, R. V., & Schnabel, P. (2001). Same-sex sexual behavior and psychiatric disorders: Findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Archives of General Psychiatry, 58,* 85–91.

Sell, R. (2007). Defining and measuring sexual orientation for research. In I. H. Meyer & M. E. Northridge (Eds.), *The health of sexual minorities: Public health perspectives on lesbian, gay, bisexual and transgender populations* (pp. 355–374). New York: Springer.

Stueve, A., O'Donelle, L. N., Duran, R., San Doval, A., & Blome, J. (2001). Time-space sampling in minority communities: Results with young Latino men who have sex with men. *American Journal of Public Health, 91,* 922–926.

Sudman, S. (1976). *Applied sampling*. New York: Academic Press.

Wilson, P. A., Valera, P., Ventuneac, A., Balan, I., Rowe, M., & Carballo-Diéguez, A. (in press). Race-based sexual stereotypes and sexual partnering behaviors of men who use the Internet to engage in bareback sex. *Journal of Sex Research.*

Winkelstein, W., Lyman, D. M., Padian, N., Grant, R., Samuel, M., & Wiley. J. A. (1987). Sexual practices and risk of infection by the human immunodeficiency virus: The San Francisco Men's Study. *Journal of the American Medical Association, 257,* 321–325.

Xia, Q., Tholandi, M., Osmond, D. H., Pollack, L. M., Zhou, W., Ruiz, J. D., et al. (2006). The effect of venue sampling on estimates of HIV prevalence and sexual risk behaviors in men who have sex with men. *Sexual Transmitted Diseases, 33,* 545–550.

Received October 23, 2008
Revision received October 23, 2008
Accepted November 4, 2008 ■