

Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations*

DOUGLAS D. HECKATHORN, *University of Connecticut*

A population is "hidden" when no sampling frame exists and public acknowledgment of membership in the population is potentially threatening. Accessing such populations is difficult because standard probability sampling methods produce low response rates and responses that lack candor. Existing procedures for sampling these populations, including snowball and other chain-referral samples, the key-informant approach, and targeted sampling, introduce well-documented biases into their samples. This paper introduces a new variant of chain-referral sampling, respondent-driven sampling, that employs a dual system of structured incentives to overcome some of the deficiencies of such samples. A theoretic analysis, drawing on both Markov-chain theory and the theory of biased networks, shows that this procedure can reduce the biases generally associated with chain-referral methods. The analysis includes a proof showing that even though sampling begins with an arbitrarily chosen set of initial subjects, as do most chain-referral samples, the composition of the ultimate sample is wholly independent of those initial subjects. The analysis also includes a theoretic specification of the conditions under which the procedure yields unbiased samples. Empirical results, based on surveys of 277 active drug injectors in Connecticut, support these conclusions. Finally, the conclusion discusses how respondent-driven sampling can improve both network sampling and ethnographic investigation.

Introduction

"Hidden populations" have two characteristics: first, no sampling frame exists, so the size and boundaries of the population are unknown; and second, there exist strong privacy concerns, because membership involves stigmatized or illegal behavior, leading individuals to refuse to cooperate, or give unreliable answers to protect their privacy. Traditional methods, such as household surveys, cannot produce reliable samples, and they are inefficient, because most hidden populations are rare. Examples include many groups at risk of contracting HIV, including active injection drug users (IDUs). Identifying these groups is crucial to the development of effective AIDS prevention interventions, and finding valid, reliable ways of sampling them is essential to evaluating interventions.

Three methods dominate studies of hidden populations: snowball sampling and other forms of chain referral samples, key informant sampling, and targeted sampling. The best known approach is *snowball sampling* (Goodman 1961): ideally, a randomly chosen sample serves as initial contacts, though in practice ease of access virtually always determines the initial sample; these subjects provide the names of a fixed number of other individuals who fulfill the research criteria. The researcher approaches these persons, and asks them to participate; and each subject who agrees is then asked to provide a fixed number of additional names. The researcher continues this process for as many stages as desired.

* An earlier version of this paper was presented at the *XI International Conference on AIDS*, Vancouver, B.C., July 1996. I thank Robert Broadhead, with whom the peer-driven intervention to combat AIDS was invented, and Denise Anthony, who conducted much of the data analysis. In addition, Dean Behrens, Robert Mills, Beth Jacobson, Boris Sergeev, David Weakliem, Jeroen Weesie and anonymous reviewers provided helpful comments and advice. Finally, I thank the *National Institute on Drug Abuse* (Grant #R01 DA08014) for its support of this research.

Erickson (1979:299) describes a number of problems afflicting snowball sampling and other chain-referral samples. First, "inferences about individuals must rely mainly on the initial sample, since additional individuals found by tracing chains are never found randomly or even with known biases." This issue is especially grave, because in the contexts where chain-referral methods are used, the initial sample usually cannot be drawn randomly. Second, chain-referral samples tend to be biased toward the more cooperative subjects who agree to participate; this problem is aggravated when the initial subjects are volunteers, because in terms of cooperation they are outliers. Third, these samples may be biased because of "masking," that is, protecting friends by not referring them, an important problem when a population has strong privacy concerns. Fourth, referrals occur through network links, so subjects with larger personal networks will be oversampled, and relative isolates will be excluded. Because of these potential biases, snowball samples typically are seen as "convenience samples" that lack any valid claim to produce unbiased and consistent samples. Still, Erickson is optimistic about the potential for resolving these problems, concluding that: "the problems in making inferences about individuals and about chains . . . are . . . in principle solvable: One can build in added incentives" (1979:299). Unfortunately, she did not develop this point.

Two additional methods have developed to overcome the difficulties afflicting snowball samples. *Key informant sampling* (Deaux and Callaghan 1985) is designed to overcome response biases by selecting especially knowledgeable respondents and asking them about others' behavior, rather than their own. For example, one might ask social workers, drug abuse counselors, public health officials, or natural opinion leaders to report on patterns of drug use and sexual behavior. This method reduces the tendency to exaggerate socially acceptable behavior and understate disreputable behavior, however, it adds several sources of bias. First, when professionals are key informants, their professional orientation may bias their responses; e.g., substance abuse counselors may exaggerate their clients' difficulties, and natural opinion leaders may feel obligated to present either an idealized or a disparaging view of those whom they influence. Second, key informants may lack sufficiently detailed knowledge; e.g., with the exception of sexual partners, knowledge about frequency of condom use tends to be strictly personal. Third, key informants do not interact with a random group of potential clients. If the key informant is a professional, the bias is a form of "institutional bias" characteristic of samples drawn from institutionalized populations. For example, just as imprisoned drug users are not randomly selected, neither are those who come into contact with drug-abuse counselors or those who are part of the entourage of a natural opinion leader. Hence the key informant approach has limitations: it introduces new sources of potential response bias, it cannot be used to access highly detailed and personal information; and the sampling may have an institutional bias.

Targeted sampling (Watters and Biernacki 1989) is a widely employed response to the deficiencies of chain-referral models. It involves two basic steps: first, field researchers map a target population (to the extent that they succeed in penetrating the local networks linking potential respondents, this prevents the under-sampling that traditional approaches would produce); and second, field researchers recruit a pre-specified number of subjects at sites identified by the ethnographic mapping, ensuring that subjects from different areas and subgroups will appear in the final sample. The adequacy of targeted sampling depends on the accuracy and comprehensiveness of the ethnographic mapping. If ethnographic mapping is near perfect, targeted sampling reduces to a form of location sampling. For example, based on an exhaustive ethnographic analysis, one could weight drug scenes based on their intensity of use by each category of user at each time of the day, and sample accordingly. Unfortunately, this is never possible because drug scenes are neither discrete nor public; while some drug copping (i.e., selling) occurs in concentrated areas, much is dispersed, occurring in private apartments and other nonpublic settings. Even an accomplished ethnographer requires

months to penetrate a small subset of such settings within one urban neighborhood. Ethnography and therefore target sampling is always limited, i.e., by the effects of the time of day when researchers recruit, where they do their recruiting, and the recruitment strategies they use (Watters and Biernacki 1989:424-426). For instance, if researchers recruit during normal business hours, they will have difficulty recruiting gainfully employed subjects. If researchers focus their activities in "obvious" locations, they introduce a subtler version of the "institutional" bias targeted sampling seeks to avoid — they oversample the most visible potential subjects, and under-sample (or miss altogether) those in less obvious niches. Researchers' own trepidations about venturing into areas where they do not feel safe further limit the diversity of the population sampled. Thus targeted sampling introduces biases that correspond to the limits of the ethnography upon which it is based.

Increasing recognition of targeted sampling's limitations and the absence of any fundamentally new sampling methods have produced renewed interest in snowball and other chain-referral methods. These methods have great potential power, as studies of social networks reveal. In populations as large as the United States, every member is indirectly associated with every other member through approximately six intermediaries (Killworth and Bernard 1978/79). This means that even the most socially isolated individuals can be reached by the sixth wave of a referral chain beginning with any arbitrarily chosen individual. Of course, realizing even a small portion of this potential requires a highly robust recruitment process, requiring, for example, that participants recruit everyone they know, and in the case of sociometric stars, this might involve many thousands of persons.

Refinements of chain-referral sampling includes Frank and Snijders' (1994) method for estimating the size of hidden populations using a one-wave snowball sample. They select a diverse set of initial subjects, each of whom then lists all members of the target population that they know. The size of the hidden population is then estimated based on the amount of overlap among the members listed. Similarly, Klovdahl (1989) proposes a "random walk" approach for analyzing network structure, using a snowball sample in which each wave consists of only one subject; its results reveal structural features of the network connecting members of the hidden population. In addition, Spreen and Zwaagstra (1994) propose a combination of snowball and targeted sampling, termed "targeted personal network sampling," which uses ethnographic mapping to locate an initial sample, which becomes the basis for network sampling. They used this approach to analyze the structure of cocaine users' personal networks.

Despite the sophistication of these extensions of chain-referral sampling, some problems remain unresolved. In particular: *"The central question in the methodological discussion about sampling and analyzing hidden populations is, basically: How to draw a random (initial) sample"* (Spreen 1992:49). This question is crucial, because it generally has been assumed that however many waves the chain-referral sampling may contain, it necessarily must reflect the biases in the initial sample.

This paper describes a new form of chain-referral sampling termed "Respondent-Driven Sampling" (RDS). It was developed as part of an AIDS prevention intervention, the *Eastern Connecticut Health Outreach (ECHO)* project, which targets active injection drug users for interviews, AIDS prevention education, and HIV testing and counseling (Broadhead and Heckathorn 1994; Broadhead et al. 1995; Heckathorn and Broadhead 1996). This process requires approximately 1 1/2 hours. RDS plays a dual role in the project. It is used both to recruit subjects into the AIDS prevention intervention, and to sample the population of active injectors. Based on an analysis drawing on Markov chains and the theory of biased networks, I show that suitable incentives can reduce the biases of chain-referral samples. Specifically, RDS produces samples that are *independent of the initial subjects from which sampling begins*. As a result, it does not matter whether the initial sample is drawn randomly. In addition, RDS

reduces the biases resulting from voluntarism and masking, and provides means for controlling the biases resulting from differences in the sizes of personal networks. Thus it resolves, or provides means for resolving, the principal problems affecting chain-referral samples.

Respondent-Driven Sampling

A principle underlying respondent-driven sampling derives from studies of incentive systems (Heckathorn 1990, 1993, 1996). Behavioral compliance can arise from two theoretically distinguishable sources. First, it can arise from individual-sanction-based control. Here an agent, such as a teacher, parent, neighbor, or AIDS prevention counselor, targets an individual for control, for example, by promising a reward for a respondent undergoing an interview. The result is a dyadic relation of the sort presumed in most analyses of influence relations, a *primary incentive*. Second, compliance can arise from *group-mediated social control* (Heckathorn 1990). For example, respondents can be rewarded not only for their own participation in a study, but also for participation they elicit from a peer. In these cases, control involves a two-step process: one or more members of the actor's group is promised a reward or threatened with a punishment based on whether the actor complies, and members of the group respond to that secondary incentive by controlling the actor. In this way, the agent's influence is amplified through the target's group. Thus, this form of control is "group mediated," and the incentives that trigger it are "secondary incentives."

A central conclusion from recent research on incentive systems (Heckathorn 1990, 1996) is that secondary incentives can be more efficient and effective than primary incentives in contexts where intragroup control is cheap and effective, as when social approval is an important sanction in peer groups. A difference between primary and secondary incentives lies in the distribution of compliance costs. In the case of individualized rewards, compliance costs are internal; the targeted actor either complies or refuses, thereby bearing whatever costs are involved, and the choice is strictly personal. In contrast, in the case of secondary incentives, compliance costs are external; the targeted actor seeks to induce others to comply, and therefore others bear the compliance costs. It is usually easier to tell others to comply than to do so oneself. This difference is crucial in the case of recruitment into a study. When only primary incentives motivate participation, individuals make their own autonomous decisions about whether to cooperate. However, if secondary rewards also motivate recruiting others into the study, peers' social influence is harnessed on behalf of the sampling process.

A second reason for the potential effectiveness of secondary incentives concerns monitoring. Agents employing primary incentives as means of control must be able to monitor compliance. But police, teachers, and drug abuse counselors typically can observe only a small portion of behavior, so monitoring is difficult when activities cannot be geographically confined. In contrast, secondary incentives operate through peer influence, and peers tend to be far more effective monitors of behavior (Heckathorn 1990). When sampling hidden populations, help from peers is often the only way that many individuals can be located. For example, no one knows better who a community's IDUs are than other IDUs.

A third reason for the potential effectiveness of secondary incentives concerns differentials in response to material incentives. Some individuals may not respond to material incentives, e.g., affluent subjects may not need the money they would gain by being interviewed. In sampling based on primary incentives, if the researcher's ability to reward symbolically is limited, he or she must rely on material rewards. Generally, social researchers have little choice but to rely upon material rewards, because, except in the case of lengthy participant observation studies, they cannot develop meaningful relationships with subjects. Secondary incentives harness peer pressure, applying non-material rewards such as peer approval to secure compliance. In essence, secondary incentives convert material incentives into peer-

based symbolic incentives. For example, individuals who are too affluent to care about material rewards may defer to social pressure from less affluent or more materially-oriented peers, such as main connections or dealers.

Like other chain-referral methods, RDS assumes that those best able to access members of hidden populations are their own peers. It differs from traditional snowball sampling in two respects. First, whereas snowball sampling typically involves an incentive for participation, RDS involves a dual incentive system — the reward for being interviewed (a primary reward) plus a reward for recruiting others into the study (a secondary reward). This study also used a mix of material (monetary) and symbolic (the opportunity to help protect oneself and one's peers from a deadly epidemic) rewards. These rewards fostered a robust recruitment, in which a few initial subjects each produced chain-referral systems that yielded a large number of recruits over the course of successive waves. For example, Figure 1 depicts the largest recruitment network from Site 2, in which a single subject generated more than one hundred recruits of diverse race, ethnicity, gender, and place of residence.

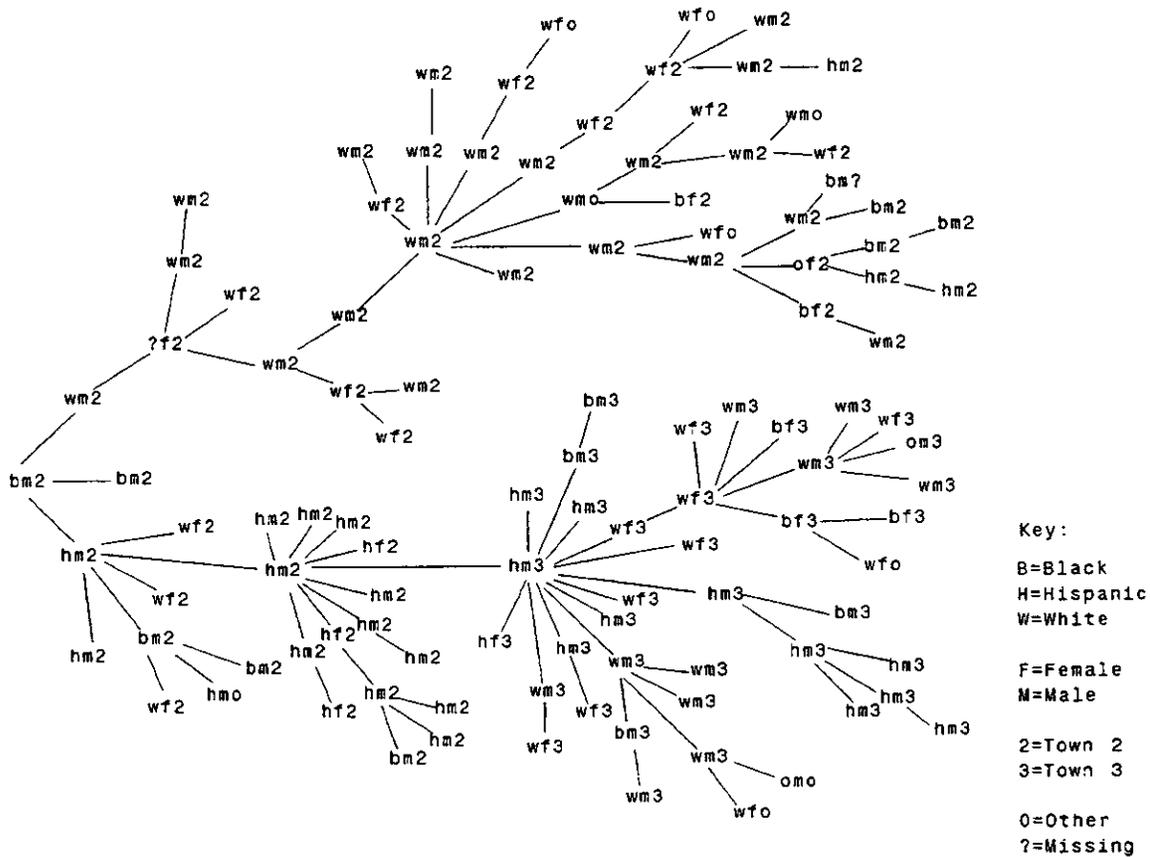


Figure 1 • Recruitment network in a respondent-driven sample, beginning from a single "seed."

A second difference between RDS and typical snowball sampling is that subjects are not asked to identify their peers to the investigator, but to recruit them into the study. This distinction is crucial when dealing with hidden populations that are subjected to considerable repression. The Netherlands are relatively tolerant, so it is not surprising that when Frank and Snijders (1994:62) asked heroin addicts to list other heroin addicts, the typical respondent gave up more than nine names. This approach would be more problematic in the

United States under current "war-on-drugs" conditions, where asking respondents to produce such lists is highly threatening, and violates the informal norm against "snitching." The RDS approach reduces masking, because it gives respondents the opportunity to allow peers to decide for themselves whether they wish to participate. For example, a subject who would not feel comfortable giving a researcher the name of a peer, may nonetheless succeed in recruiting that peer. Also, the problem of recruiting only the most cooperative subjects is reduced by combining primary and secondary incentives. Individuals who resist field researchers' appeals may nonetheless yield to appeals from peers motivated, at least in part, by secondary incentives. Even the least cooperative individuals are not immune to social pressure. Finally, it typically is assumed that chain-referral samples are biased toward subjects with large personal networks. However, an empirical test of this hypothesis did not find evidence of this (Welch 1975). Similarly, in the current study, the association between personal network size, as measured by a question that asked how many injectors the respondent knew by name or face, lacked any statistically significant association with the number of subjects recruited into the study. This may result from the recruitment quotas built into this study. Even those subjects with small personal networks could generally find at least a few subjects to recruit. Recruitment quotas reduced the ability of subjects with large personal networks to recruit more extensively than subjects with smaller networks.

The potential bias of oversampling subjects with larger personal networks can be addressed in several ways: (a) samples can be weighted inversely with subjects' self-reported network size; (b) sampling can focus on saturating targeted areas and thereby capture subjects with the full array of network sizes; or (c) special incentives can be employed to increase recruitment of subjects with traits associated with small personal networks.

Respondent-driven sampling was implemented in the following way:

- (1) Research staff recruit a handful of subjects who serve as "seeds."
- (2) Seeds are offered financial incentives to recruit their peers into the same interview they have just completed. Specifically, seeds are given several *recruitment coupons* and told that if they pass the coupons on to peers who come to the storefront for an interview, the recruiter will be paid \$10 for each recruited peer.
- (3) All new recruits are offered the same dual incentives as were the seeds. Everyone is rewarded both for completing the interview, and for recruiting their peers into the research. Given adequate incentives, this mechanism creates an expanding system of chain-referrals in which subjects recruit more subjects, who recruit still more subjects, and so on from wave to wave. To ensure that a broad array of subjects have an opportunity to recruit, to prevent the emergence of semi-professional recruiters, and to preclude turf battles over recruitment rights, each subject was limited to three initial coupons. Sets of three additional coupons were given only under two conditions, that three subjects had been successfully recruited, and that they had been effectively educated about HIV prevention issues as measured using a structured assessment test. A by-product of this recruitment quota was to increase the number of waves of recruitment required to saturate the population. Given the combination of incentives for recruitment and education, the average cost per recruited subject was about \$14.
- (4) The trait defining membership in the population must be objectively verifiable, lest respondents react to the recruitment incentives by enlisting persons who are not part of the hidden population (e.g., noninjectors). In this study, verification was accomplished through a seven-step screening protocol used to confirm injection drug use. For example, the first step was to look for recent track marks. If marks were not found, subsequent steps required the subject to demonstrate detailed acquaintance with drug preparation and injection techniques.

- (5) Subject duplication occurs when a subject seeks to participate in a study under multiple identities, and subject impersonation occurs when a subject pretends to be another, perhaps as a means to collect the latter's reward for recruitment. These potential problems were overcome using a subject identification database recording subjects' identifying physical characteristics, including gender, age, ethnicity, height, weight, scars, tattoos, and some biometric measures.
- (6) Targeting specific subgroups within a population can occur through the use of "steering" incentives, that is, extra bonuses paid for recruiting specific categories of subjects. For example, a modest (\$5) bonus was offered for recruiting female injectors.
- (7) Sampling can be ended either when the targeted community is saturated, or when a minimum target sample size has been reached and the sample composition has reached a stable composition with respect to the traits upon which the research focuses.

These procedures have inherent limitations. Like any chain-referral sampling procedure, RDS is suitable only for sampling populations with a contact pattern; the activities that constitute membership in the population must create connections among population members, as when drug users purchase or share drugs, or when high-risk sexual activities take place. Therefore, this method is not suitable for drawing national samples. The size of the area within which sampling can be effective depends on the contact pattern's geographic extensiveness, which in turn depend on the availability of transportation to the respondents. Nor is RDS suitable for sampling hidden populations whose members share no ties. Second, the trait defining membership in the population must be verified objectively, lest interview and recruitment incentives motivate some respondents to claim falsely that they are members of the population. This is not a problem unique to RDS; it occurs whenever interview respondents are rewarded, but the recruitment incentive may aggravate the problem, requiring a well-tested screening protocol.

The Setting

The RDS was implemented in two small cities in Connecticut with populations of 28,500 and 42,800, and substantial injection drug use and associated AIDS cases. As of July 1996, 185 cases of AIDS had been diagnosed in the first site, and 89 at the second, with another 116 cases in a nearby town of 59,500 that also fell within the sampling area. About one half of the cases in these towns — 55%, 49% and 47% respectively — involved drug injection. Unemployment among the IDU respondents was high, i.e., 71.7% and 78.9% at sites 1 and 2 respectively. Sampling at the two sites was sequential. It began at the first site in March 1994, continuing until March 1995. Sampling at the second site began in March 1995 and the data reported here extend through March 1996. Therefore, I report on exactly one year of operation at each site. Both projects operated out of storefronts located in active drug using scenes. A program staff of three health educators scheduled appointments for interviews during three weekdays.

A Formal Model

When viewed analytically, a RDS creates a stochastic process in which each recruiter's social characteristics affect the characteristics of the recruits. In the case of race and ethnicity, this means that recruiters of each ethnic group generate a distinct ethnic mix of recruits. For example, from Table Ia it is apparent that in Site 1, non-Hispanic whites, on average, recruited 70% whites, 19% non-Hispanic blacks, 10% Hispanics, and 2% others. Similarly, non-Hispanic blacks recruit, on average, 28% whites, 56% blacks, 6% Hispanics, and 9% others, so the tendency toward in-group recruitment persists. This pattern continues among

Hispanics, whose in-group recruitment rate is 59%. Hence, recruiting occurs preponderantly within the ethnic group, but cross-ethnic recruitment is also common.

Table I • Characteristics of Recruits, by Characteristics of Recruiter, Site 1

Table Ia • Recruitment by Ethnicity

<i>Ethnicity of Recruiter</i>	<i>Race/Ethnicity of Recruit</i>				<i>Total</i>
	<i>W</i>	<i>B</i>	<i>H</i>	<i>O</i>	
Non-Hispanic White (W)	69.8%	19%	9.5%	1.6%	100% (63)
Non-Hispanic Black (B)	28.1%	56.3%	6.3%	9.4%	100% (32)
Hispanic (H)	23.5%	5.9%	58.8%	11.8%	100% (17)
Other (O)	50%	50%	0%	0%	100% (4)
Total Distribution of Recruits	50.9% (59)	28.4% (33)	15.5% (18)	5.2% (6)	100% (116)
Equilibrium	49%	29.7%	15.8%	5.4%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = .91% ($r = .998$)					

Table Ib • Recruitment by Gender

<i>Gender of Recruiter</i>	<i>Female</i>	<i>Male</i>	<i>Total</i>
Female	29.7%	70.3%	100% (37)
Male	23.9%	76.1%	100% (88)
Total Distribution of Recruits	25.6% (32)	74.4% (93)	100% (125)
Equilibrium	25.4%	74.6%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = .2%			

Table Ib reports the corresponding data for gender. No in-group recruitment pattern is apparent in this case, because both females and males recruit between 70% and 76% males. This may reflect the male-dominated character of injection drug scenes. However, because rewards for recruiting females were greater than those for recruiting males, their expected effect was to increase in-group selection among women, strengthening their incentives to recruit one another, while reducing in-group selection among men, because their incentive was to recruit women. This shows how steering incentives can alter tendencies toward in-group or out-group selection.

An intermediate level of in-group selection is apparent in Table Ic's report of recruitment based on drug preference. Most subjects in the study report that heroin is their drug of first choice, so drug preference was dichotomized between heroin, and an "other" category that includes cocaine and crack, methadone, cannabis, speedballs (a combination of cocaine and heroin), and alcohol. A tendency for subjects to recruit from their own group is apparent, but it is weaker than the effects of ethnicity. Finally, Table Id reports recruitment based on

Table Ic • Recruitment by Drug Preference

<i>Drug Preference of Recruiter</i>	<i>Drug Preference of Recruit</i>		
	<i>Heroin</i>	<i>Other</i>	<i>Total</i>
Heroin	87.7%	12.3%	100% (81)
Other	67.6%	32.4%	100% (34)
Total Distribution of Recruits	80% (94)	20% (21)	100% (105)
Equilibrium	84.6%	15.4%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = 2.86%			

Table Id • Recruitment by Location

<i>Location of Recruiter</i>	<i>Location of Recruit</i>			<i>Total</i>
	<i>In Town (Town 1)</i>	<i>In Area</i>	<i>Out of Area</i>	
In Town (Town 1)	84.3%	8.4%	7.2%	100% (83)
In Area	60%	10%	30%	100% (10)
Out of Area	50%	8.8%	41.2%	100% (34)
Total Distribution of Recruits	73.2% (93)	8.7% (11)	18.1% (23)	100% (127)
Equilibrium	77.5%	8.6%	13.9%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = 2.86% ($r = .997$)				

area of residence. Here the tendency toward in-group recruitment is variable: whereas residents of Site 1's town recruit internally 84% of the time, subjects from the area (i.e., from contiguous towns) or out of the area (i.e., from more distant towns) recruit strongly both from within and outside of town. This reflects the town's position as a regional drug distribution center. IDUs travel considerable distances to purchase drugs in the town, and in the process they develop network connections both in the town and in their area of origin.

Sampling as a Markov Process

RDS recruitment has two important characteristics. First, there are a limited number of states (e.g., types of ethnicity) that subjects can assume. Second, any subject's recruits are a function of his or her *type*, such as his or her ethnicity; and not of previous events, such as who recruited the recruiter. This requirement is satisfied in the case of Table Ia's data because there is no significant association between the ethnicity of a recruiter's recruits and the

ethnicity of the recruiter's recruiter.¹ Hence, recruitment is a memoryless process. For example, white recruiters who were themselves recruited by whites, recruited about the same mix of subjects as did white recruiters who were recruited by Hispanics or blacks. Therefore, RDS recruitment in this case qualifies as a *first-order Markov process*,² and can be represented in the form of the network depicted in Figure 2a. The recruitment process can be conceptualized as a point moving from node to node in this network, each representing a distinct state of the system. At any instant, the current location of the point indicates the most recent recruit's ethnicity. The arrows leaving that point to other nodes indicate the ethnicity of the next recruit, and the number associated with each arrow indicates the probability that this path will be taken.

Markov processes are of several basic types. Some have absorbing states, such that when one enters that state, no exit is possible. This occurs when an ethnic group is totally isolated, such that its members only recruit one another. Figure 2a, indicates this does not fit this network; there is a substantial in-group selection bias, but it is less than total. Indeed, following the arrows in Figure 2a's network, one can reach any point in the network from any other point. Therefore, this network is "ergodic" (Fararo 1973:280). In contrast, Figure 2b illustrates a nonergodic network that will be discussed below, in which the rightmost two nodes are absorbing states. Returning to Figure 2a, the network is non-cyclic, in that any node can be reached during any time period (e.g., odd versus even time periods). Because the Markov process is both ergodic and non-cyclic, Figure 2a's network is termed "regular."

Two theorems regarding regular Markov processes are relevant to understanding RDS. First, the "law of large numbers for regular Markov chains" (Kemeny and Snell 1960:73) states that the probability that a system will be in any given state over the course of a large number of steps is independent of its starting state. The implication for RDS recruitment is that:

THEOREM ONE: As the recruitment process continues from wave to wave, an equilibrium mix of recruits will eventually be attained that is independent of the characteristics of the subject or set of subjects from which recruitment began.

Thus, if recruitment operates until equilibrium is reached, and the resulting recruitment network qualifies as a regular Markov process, it avoids the central problem for sampling hidden populations — that the sample's characteristics merely reflect the initial sample. Instead, a respondent-driven sample is wholly independent of the initial set of subjects. This point is illustrated graphically in Figures 3a and 3b, in which Table 1a is used as the basis for two alternative simulations. Figure 3a draws on Table 1a's data to mathematically project what would have happened had recruitment begun with only non-Hispanic African-American seeds. After the first wave, those subjects would have on average recruited 28.1% whites, 56.3% blacks, 6.3% Hispanics, and 9.4% others, so blacks would predominate. After the second wave, the composition of recruits would change, because 28.1% of recruiters would be white, and they would recruit, on average, 69.8% whites, 19% blacks, 9.5% Hispanics, and 1.6% others (i.e., Native Americans and Asian Americans). Similarly, 6.3% of second-wave recruiters would be Hispanic, and they would recruit on average 23.5% whites, 5.9% blacks, 58.8% Hispanics, and 11.8% others. Combining these with the recruits of the blacks and others, the overall composition of the second-wave recruits is 40% whites, 43.7% blacks, 9.9% Hispanics, and 9.5% others. As Figure 3a illustrates, this trend continues, from wave to

1. A series of chi-square analyses was conducted to determine whether the ethnicity of a recruiter's recruiter significantly affected the ethnic composition of the recruiter's recruits. For example, white recruiters were differentiated based on whether they were recruited by a Hispanic, black, white, or an "other" person, and the ethnic composition of their recruits were compared. No significant differences were found, e.g., in the case of white recruiters, the significance level was 0.44.

2. A Markov chain is formally equivalent to contingency tables that display changes in state from one period of time to another, and as such they can be analyzed using log-linear models (see Bishop et al. 1975:257-279).

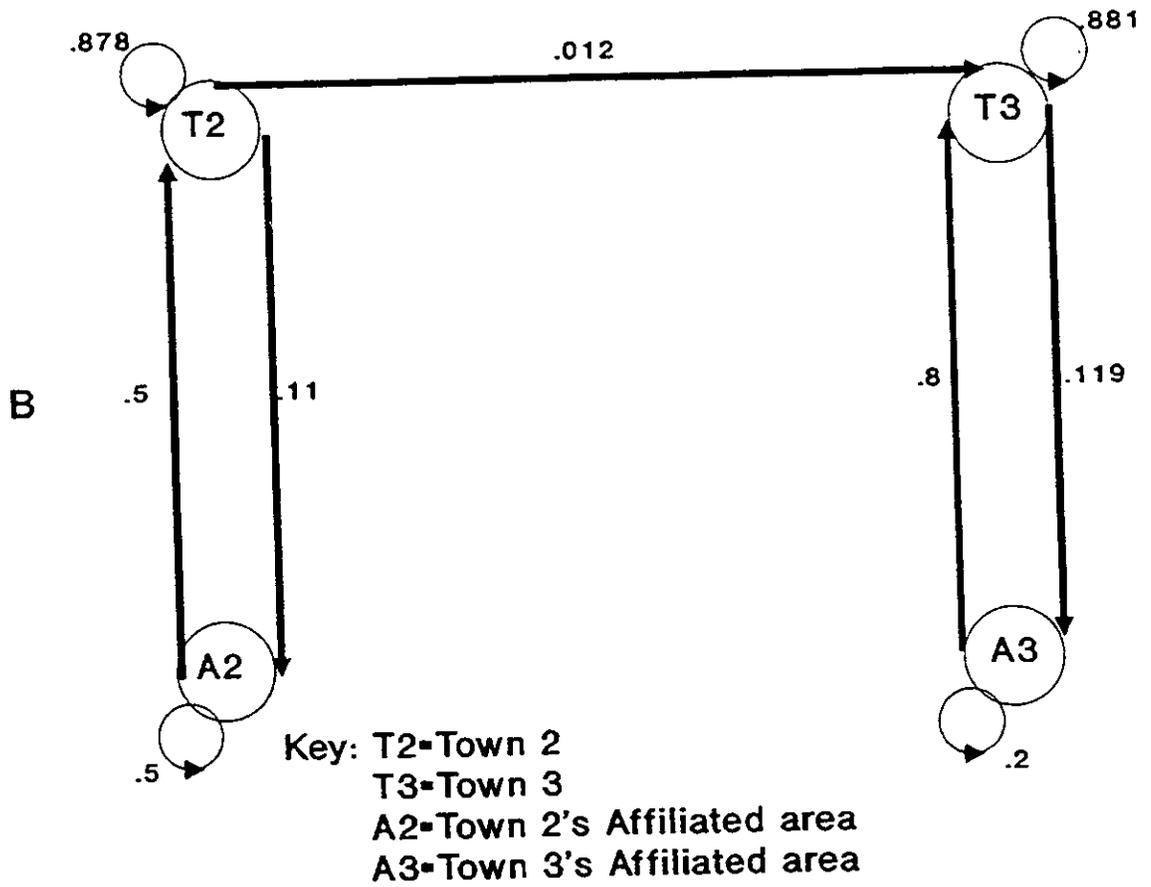
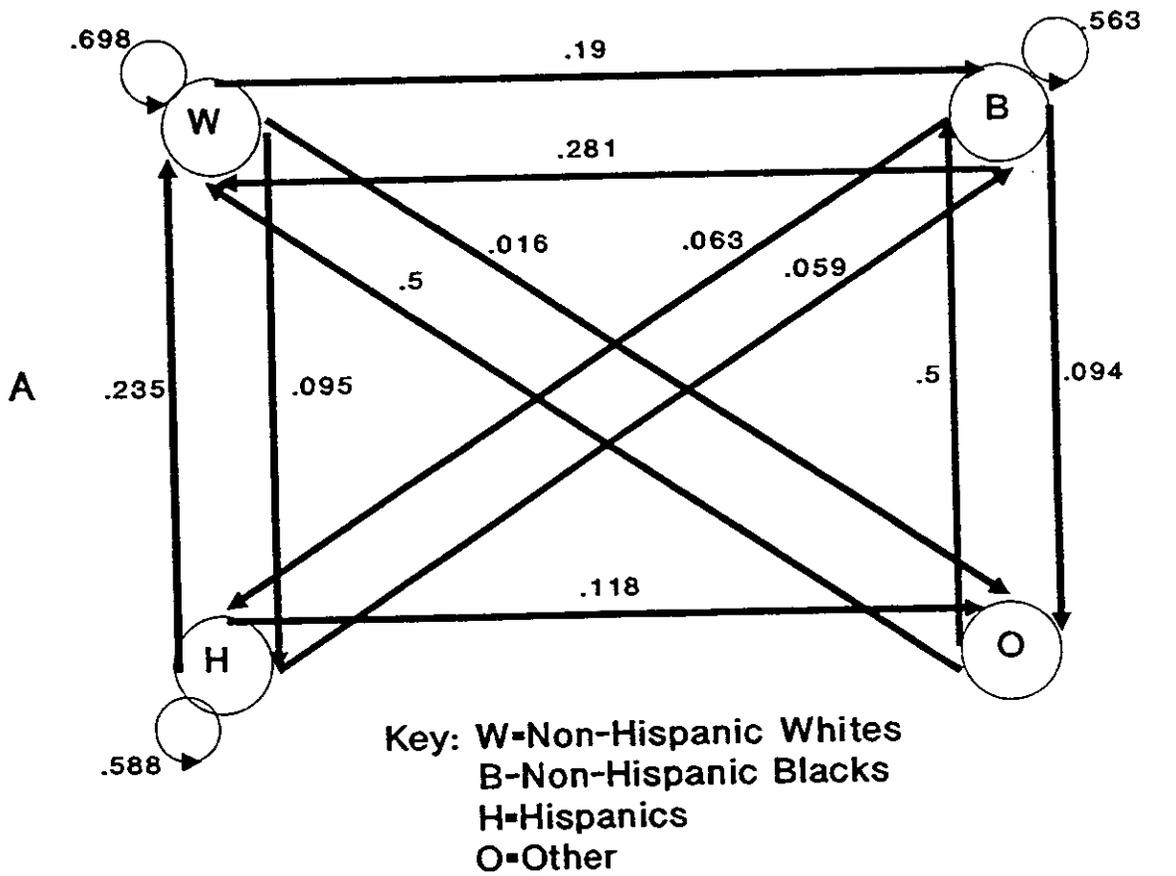


Figure 2 • Markov chain networks, where nodes represent subject characteristics, and arrows represent recruitment probabilities

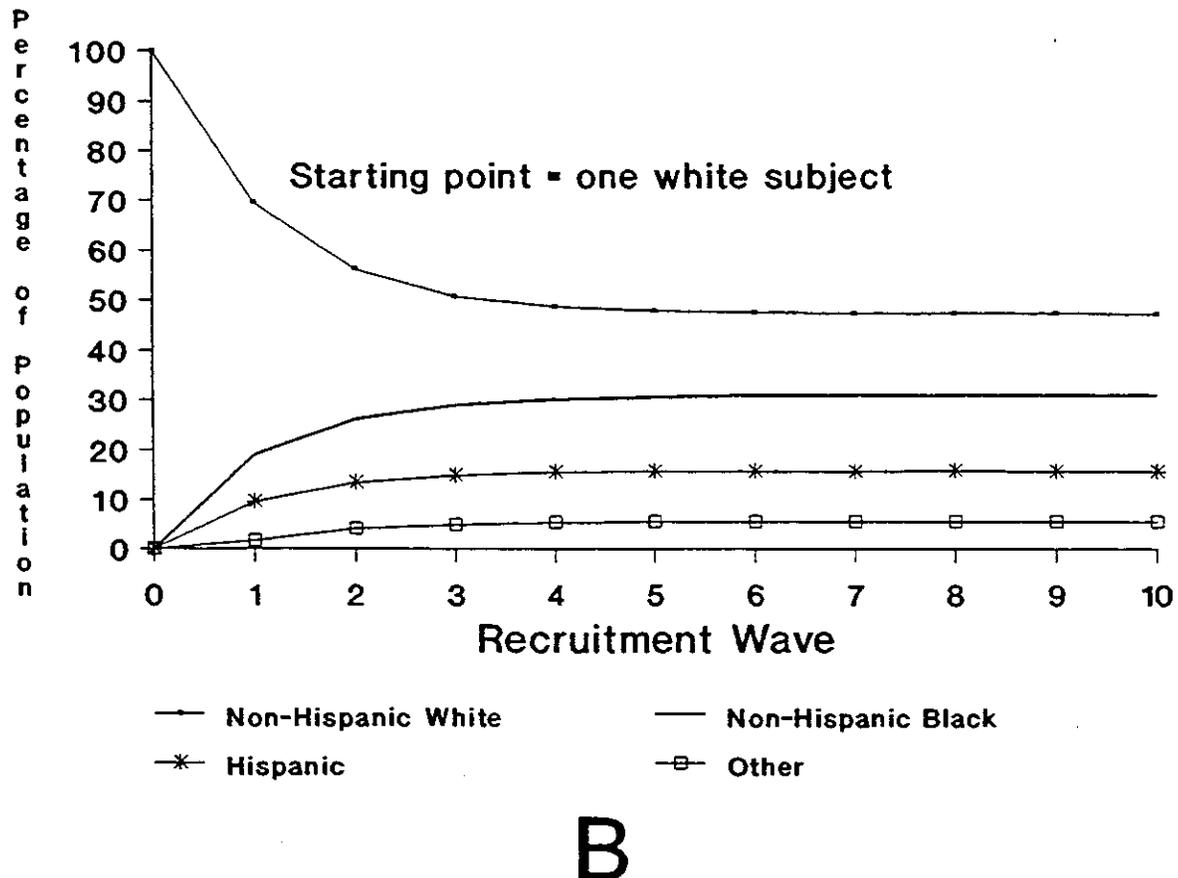
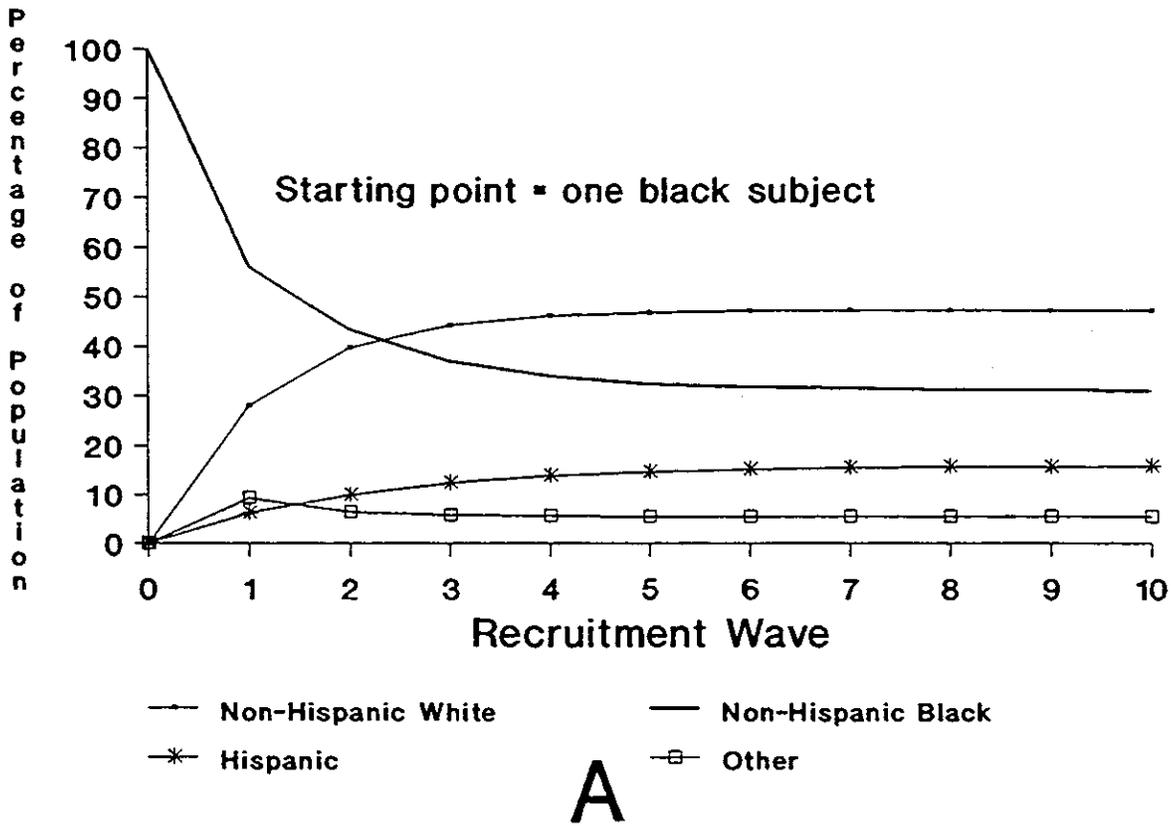


Figure 3 • Race and Ethnicity of Recruits in a Respondent-Driven Sample, Beginning with a Single Black or White Subject

wave, until recruitment stabilizes, so the proportion of blacks declines to approximate their occurrence in the sample. Similarly, Figure 3b shows what would have happened had recruitment begun with one or more white seeds. The initial over-sampling of whites would decline from wave to wave, until an equilibrium recruitment pattern was attained. Note that, consistent with Theorem One, the ultimate recruitment pattern is independent of the ethnicity of the initial subject(s) with which recruitment began. That is, the ultimate recruitment pattern is identical, whether the initial seed(s) are black or white. The weakness of chain-referral samples — that the ultimate sample depends on the choice of initial subjects — is not inherent in this form of sampling. Instead, this bias arises only when sampling does not continue through enough waves for equilibrium to be reached.

This conclusion applies generally to snowball samples, if two conditions are met. First, the snowball process must continue to the point of equilibrium. This requires more stages than occur in most snowball samples. Second, the recruitment process must correspond to a regular Markov process. This constraint is less severe than may be at first apparent. For example, a sample corresponding not to a first-order but to a higher order Markov process does not produce severe problems. Its effect may be merely to slow the approach to equilibrium, so sampling may require additional waves.

The law of large numbers for regular Markov chains provides a means for computing analytically the equilibrium sampling of groups (see Kemeny and Snell 1960:72). The equilibrium state (E) of Table Ia's system is found by solving the following system of equations:

$$1 = E_w + E_b + E_h + E_o \quad (1)$$

$$E_w = .698 E_w + .281 E_b + .235 E_h + .5 E_o \quad (2)$$

$$E_b = .19 E_w + .563 E_b + .059 E_h + .5 E_o \quad (3)$$

$$E_h = .095 E_w + .063 E_b + .588 E_h \quad (4)$$

where E_w , E_b , E_h , and E_o are the equilibrium sample proportions for non-Hispanic whites, non-Hispanic blacks, Hispanics, and other groups, respectively. Equation one expresses the fact that the sum of proportional distributions of population must sum to one. The subsequent equations express each group's equilibrium size as a function of the equilibrium sizes of all members of the group, and each group's proportional recruitment of each group. Because this is a system of four linear equations with four unknowns, there is a unique solution, that is, $E_w = .490$, $E_b = .297$, $E_h = .158$, and $E_o = .054$.³

A matter of great significance in determining the practical import of the above theorem is the rate at which the equilibrium is reached. That this convergence ultimately would occur would not matter, if convergence were slow. Fortunately, regular Markov chains are characterized by what Kemeny and Snell (1960:72) describe as "a very fast kind of convergence." This conclusion is based on a theorem that proves that convergence occurs geometrically. The implication is that:

THEOREM TWO: The subject pool generated by a respondent-driven sample approaches equilibrium at a rapid (i.e., geometric) rate.

The implication, except in a special class of cases discussed below, is that convergence occurs within a handful of recruitment waves. This conclusion is consistent with the above discussed data. For example, the mean absolute discrepancy between the actual subject composition by race/ethnicity and the computed equilibrium composition is less than 1%, showing that equilibrium has been approximated. Similarly, recruitment based on gender, drug preference, and recruitment location (see Tables Ib to Id) also closely approximates equilibrium,

3. A program is available from the author that computes these equilibria for systems with up to ten groups. It also computes the composition of the sample, as it changes from wave to wave, based on any specification of initial subjects, and indicates the number of waves required for equilibrium to be reached. For a copy, send a blank disk, along with a stamped, self-addressed diskette mailer to the author. The program is menu driven and requires an IBM compatible with VGA graphics.

as reflected in mean discrepancies between the sample and equilibrium of less than 3%. This close approximation to equilibrium is theoretically expected, in that irrespective of the initial distribution of seeds, approximating equilibrium to within a tolerance of 2% within any of Table I's matrices never requires more than six recruitment waves. A further implication of the theorem is that the greater the extent to which the initial sample approximates the equilibrium sample, the quicker will be the approach to equilibrium. Hence, choosing a diverse sample speeds the approach to equilibrium.

Table II • Recruitment by Area, Site 2

Table IIa • Recruitment by Location, Site 2

<i>Location of Recruiter</i>	<i>Location of Recruit</i>			<i>Total</i>
	<i>Town 2</i>	<i>Town 3</i>	<i>Other</i>	
Town 2 (Site 2's location)	87.8%	1.2%	11.0%	100% (82)
Town 3	0%	88.1%	11.9%	100% (59)
Other	22.2%	55.6%	22.2%	100% (9)
Total Distribution of Recruits	49.3% (74)	38.7% (58)	12.0% (18)	100% (150)
Equilibrium	23.7%	63.3%	13.0%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = 17.09% ($r = .432$)				

Table IIb • Recruitment by Location, with Partitioning of the "Other" Group, Site 2

<i>Location of Recruiter</i>	<i>Location of Recruit</i>				<i>Total</i>
	<i>Town 2</i>	<i>Town 3</i>	<i>Town 2's Affiliated area (A2)</i>	<i>Town 3's Affiliated area (A3)</i>	
Town 2 (Site 2's location)	87.8%	1.2%	11.0%	0%	100% (82)
Town 3	0%	88.1%	0%	11.9%	100% (59)
Town 2's Affiliated area (A2)	50%	0%	50%	0%	100% (4)
Town 3's Affiliated area (A3)	0%	80%	0%	20%	100% 5
Total Distribution of Recruits	49.0% (74)	38.4% (58)	7.3% (11)	5.3% (8)	100% (150)
Equilibrium	0%	87.1%	0%	12.9%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = 28.15% ($r = .330$)					

Consider now the case of Site 2, in which a more complex theoretic analysis is required to make sense of the data. Table IIa shows recruitment as a function of the recruiter and recruit's towns of origin for Site 2. Here *Town Two* is Site 2's location. The seeds were drawn from this site, because an aim of the study was to determine how geographically extensive recruitment would become. Hence no effort was made to specify the sampling universe in

Table IIc • Recruitment by Location, Town 2 and Vicinity

<i>Location of Recruiter</i>	<i>Location of Recruit</i>		<i>Total</i>
	<i>Town 2</i>	<i>Affiliated area (A2)</i>	
Town 2 (Site 2's location)	88.9%	11.1%	100% (81)
Affiliated area (A2)	50%	50%	100% (4)
Total Distribution of Recruits	87.1% (74)	12.9% (11)	100% (85)
Equilibrium	81.8%	18.2%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = 5.26%			

Table IIId • Recruitment by Location, Town 3 and Vicinity

<i>Location of Recruiter</i>	<i>Location of Recruit</i>		<i>Total</i>
	<i>Town 3</i>	<i>Affiliated area (A3)</i>	
Town 3	88.1%	11.9%	100% (59)
Affiliated area (A3)	80%	20%	100% (5)
Total Distribution of Recruits	87.5% (56)	12.5% (8)	100% (64)
Equilibrium	87.5%	12.5%	100%
Mean Discrepancy, Distribution of Recruits and Equilibrium = 0%			

advance. Unlike most studies, this served as a outcome variable. *Town Three* is a larger town about 15 miles away, and *Other* refers to a mix of locations, either in between or (in a few cases) more distant. In-group recruitment within the two towns is very strong (i.e., 88%), and cross-recruitment is almost nonexistent (i.e., 1.2% and 0%). Recruitment began from Town 2 and later extended to Town 3, a process that shows the ability of RDS to break out of one isolated group to penetrate another similarly isolated group. How this occurred is depicted in Figure 1, which shows the largest recruitment network from this site. During the third wave, 5 1/2 months into the 12 month study period, a Hispanic male subject from Town 2 recruited another Hispanic male from Town 3. From this tiny foothold grew a substantial branch of the recruitment tree (see the bottom-right quadrant of Figure 1). That subject directly or indirectly recruited 40 other subjects from Town 3 and the surrounding area. This shows vividly how a sufficiently robust chain-referral system can exploit existing network connections to expand into isolated groups.

The relative isolation of the two towns derives from transportation difficulties. Many subjects (53.1%) report that they do not have access to a car, so hourly bus service is a primary means of transportation. Most subjects traveling from Towns 3 to 2 began with 15 mile bus ride, and then walked an additional mile to the ECHO Project storefront. Though they were compensated for their bus fares, this reflects a remarkable commitment to participation in the study. That subjects continued to make this trek during the Connecticut winter is especially remarkable.

There is a sharp contrast between recruitment by area at the second and the first sites. Whereas equilibrium was closely approximated by the sample distribution in the first site

(mean discrepancy, $D=2.86\%$), the divergence is large at the second site ($D=17.09\%$). This reflects an over-sampling in Site 2 of Town 2 (49.3% versus 23.7%), and an under-sampling in Town 3 (38.7% versus 63.3%). It might seem that this discrepancy results merely because the system has not as yet reached equilibrium. This might appear to explain much of the departure from equilibrium, because at present in this on-going study, the first subject from Town 3 appeared in the sample 5 1/2 months into the 12 month study period, and the tenth wave has only just begun. Based on this incomplete movement towards equilibrium, one could then weight the sample. For example, given that Town 2 is over-sampled by 25.6% (i.e., 49.3%-23.7%), and Town 3 is under-sampled by 24.6% (i.e., 63.3%-38.7%), these towns can be assigned weights of .481 (i.e., 23.7%/49.3%) and 1.64 (i.e., 63.3%/38.7%), respectively when analyzing ethnicity, gender, and drug preference. The "Other" area had already closely approximated equilibrium, so it is assigned a virtually neutral rate of 1.08 (i.e., 13%/12%). Such an account is theoretically plausible, but a closer examination of the data shows that it does not fit this particular case.

As represented in Table IIa, the Other group serves as a bridge between Towns 2 and 3, recruiting equally (22%) from both towns. Why then in Figure 1's recruitment network did none of the 10 subjects from the "other" area serve as bridges between the two towns, i.e., why were there no Town 2/Other/ Town 3 recruitments? Given the modest number of cases, this could be a coincidence, but the same occurs in Site 2's other recruitment networks. However, on closer inspection, the Other category disaggregates into two mutually exclusive groups. First, there are persons who live outside of Town 2 and have connections only to that town and to their own locality. This may be termed the "affiliated area" of Town 2. Second, Town 3 also has an affiliated area. Because of the considerable distance between the towns, no subjects reside in intermediate areas with network connections to both towns. Therefore, the appearance that the Other category can serve as a bridge between the two towns is an illusion created by inappropriate categorization. As this case illustrates, in RDS analysis the assignment of subjects to categories is by no means neutral. It must take into account the structure of network connections. Specifically, if a category includes individuals with quite disparate network structures, they should be further differentiated by category.

Table IIb shows the matrix produced when the Other category in Table IIa is divided based on whether they are affiliated nonresidents of Towns 2 or 3. As is apparent by inspection, this does not bring the system closer to equilibrium. Indeed, the system is farther from equilibrium than previously ($D = 28.15\%$), and the content of the equilibrium has become strange, because recruitment is projected to die out wholly in Town 2 and its affiliated area. The reason for this problem is apparent from inspection of Figure 2b, which depicts Table IIb's recruitment data in network form. Note that the network is nonergodic, in that every point in the network cannot be reached directly or indirectly from every other point. This occurs because nodes for Town 3 and its affiliated area serve as absorbing states. This is why the equilibrium for this system (see Table IIb) shows recruitment as ceasing in Town 2 and its area, and continuing only in Town 3 and its area.

The solution to this problem is straightforward. Because the network is nonergodic, this system does not generate a regular Markov process. Therefore, data from this site can best be analyzed by partitioning the sample into two subsamples, one for Town 2 and its affiliates, and one for Town 3 and affiliates. Tables IIc and IId illustrate this process for Table IIb's data. Both systems reach similar equilibria, with about 87% recruitment within the town, and the remainder from affiliated outlying areas; equilibrium is closely approximated in both systems. In sum, whereas Site 2 initially appeared to contain a single respondent-driven sample with anomalous properties, it turns out on closer examination to consist of two distinct samples. This possibility should be suspected whenever, despite a substantial sample size, departures from equilibrium are great.

Assessing Bias in Respondent-Driven Samples

An ideal sampling procedure yields not only a sample independent of its starting point, but also an unbiased sample of the underlying population, with a known degree of consistency from which confidence intervals can be computed. However, because of the absence of probability samples in studies of hidden populations, a more modest goal has been to devise means for drawing samples that produce "a good cross-section of the target population," or "the coverage of heterogeneity in the target population" (Spren and Zwaagstra 1994:478). Such samples are termed "representative." An assessment of RDS requires examining both the extent to which it fulfills this modest goal, and whether it might provide a basis for fulfilling the more stringent ideal of probability sampling. Returning again to the data from Site 1, members of the major ethnic groups tend to recruit differentially from within their own ranks (see Table Ia). This reflects the social structure in which these subjects are embedded; intra-ethnic connections are more common than inter-ethnic connections. A form of network analysis, "biased network theory," has been used to analyze a variety of relationships, including marriages and friendships (Fararo and Skvoretz 1984; Rapoport 1977). The theory's essential idea is that a structured social system's social linkages will be non-random: some relationships will be more probable than others, where "biases" refer to any departures from a fully random pattern of connection. Biases may consist of either a tendency toward in-group affiliation, as typically occurs in friendships; or a tendency toward out-group affiliation, as in exogamous marriage systems.

In an unstructured group, recruitment merely reflects the prevalence of each group within the population. Every recruiter, irrespective of group identity, recruits the same mix of subjects. From inspection of Table I, that obviously does not fit the data. In a system where group affiliation affects selection, the members recruited by each individual reflect both the recruiter's biases, and the prevalence of different types of members within the population. Selection in such a system can be modeled as a process with two conditional steps (Fararo and Skvoretz 1984:233). First, an *inbreeding bias event* for a member of group X either occurs, with probability I_x , or fails to occur with probability $1-I_x$. If the inbreeding bias event occurs, selection occurs from the in-group *with certainty*. Second, if the inbreeding event does not occur, any individual is selected randomly from the system's population irrespective of group membership. Therefore, the probability of selecting a member of any group is equal to the proportion of that group's members in the system. In this formulation, selection of an out-group member proves that the inbreeding event did not occur, whereas selection of an in-group member can occur either because of the inbreeding event, or by random selection from the system's population. As thus conceptualized, the magnitude of inbreeding reflects a mix of cultural and situational factors, ranging from cultural emphasis on in-group affiliation to ease of transportation between geographically distinct groups. It can also be affected by steering incentives, such as those used to increase recruitment of female IDUs.

This model involves a host of simplifying assumptions. For example, the occurrence of inbreeding is treated as a dichotomous event, whereas gradations appear more empirically plausible. However, such simplifying assumptions are not necessarily undesirable. Formal heuristic models should never be made more complex than required to capture the fundamentals of the process they are intended to represent, lest they lose tractability and fail to yield clear conclusions. Ideally, validation of the model requires an assessment of its robustness. First, simplifying assumptions are replaced by more realistic assumptions. Then, whether the model's fundamental conclusions remain valid is assessed. The aim is to identify the simplest robust model. However, such an assessment exceeds the scope of this paper. What can be said at present is that the Fararo-Skvoretz model captures the two essential features of inbreeding, i.e., that groups vary in their strength of inbreeding, and that the

resulting selections produce the structure of in-group and out-group affiliations within the system. Hence the model is sufficient for heuristic purposes.

Consider the case of a system composed of N groups, A, B, C, \dots, N . The probability of a member of a group X selecting from the in-group, S_{xx} , is the sum of the probability of selection being controlled by inbreeding, an event with probability I_x , and the probability that inbreeding did not occur $(1 - I_x)$, weighted by the proportion of members of group X in the population, P_x , i.e.,

$$S_{xx} = I_x + (1 - I_x) P_x \tag{5}$$

By similar principles, the probability of a member of group X selecting a member of any out-group, Y , can be computed. The probability of selecting a member of group Y is the probability that inbreeding did not occur $(1 - I_x)$, weighted by the proportion of members of group Y in the population, P_y , i.e.,

$$S_{xy} = (1 - I_x) P_y \tag{6}$$

With this set of equations, the model specifies the relationship between the underlying population proportion (P), and selection probabilities (S) of the sort depicted in Table Ia.

Based on the above pair of equations, equilibrium sample size can be expressed as a function of both population and inbreeding terms. To simplify the analysis, consider first the two-subgroup case. Consistent with the law of large numbers for regular Markov chains, the equilibrium proportion of members of each group is found by solving a system of two linear equations:

$$1 = E_a + E_b \tag{7}$$

$$E_a = S_{aa} E_a + S_{ba} E_b \tag{8}$$

which yields,

$$E_a = \frac{S_{ba}}{1 - S_{aa} + S_{ba}} \tag{9}$$

This in turn can be expanded by substitution from equations 5 and 6 to express the equilibrium sample in terms of the inbreeding terms and population:

$$E_a = \frac{P_a(1 - I_b)}{1 - (I_a + P_a(1 - I_a)) + P_a(1 - I_b)} \tag{10}$$

This simplifies to:

$$E_a = \frac{P_a(I_b - 1)}{P_a I_b - P_a I_a + I_a - 1} \tag{11}$$

This expression provides a basis for deriving conclusions, grounded in both biased-network theory and Markov chains, regarding the conditions under which a RDS will produce an unbiased sample, that is, a sample in which the proportion of each group in the sample equals that group's proportion in the population.

Given that inbreeding affects recruitment, the equilibrium group composition (E) need not correspond to the true underlying population distribution (P). The extent to which members of any given group will be sampled depends on three factors: the size of the group, its tendency toward inbreeding, and the strength of inbreeding in other groups. Partial differentiation⁴ shows that a group's proportion in the sample increases, all else equal, as the

4. Partial differentiation provides a means for determining whether the relationship between two terms is consistently positive, consistently negative, or mixed. The first step is to compute the first derivative of one term with respect to the other. The direction of the relationship is then found by determining whether, given constraints on parameter values, the derivative is always positive, always negative, or mixed. For example, to determine the relationship between equilibrium sample size for a group, and that group's inbreeding bias, differentiate E_a with respect to I_a , i.e.,

$$dE_a/dI_a = \frac{P_a(1 - I_b)(1 - P_a)}{(P_a I_b + P_a I_a + I_a - 1)^2}$$

group's size increases and as its inbreeding increases. Similarly, a group's proportion decreases with increases in inbreeding by other groups. Thus, a small group potentially might appear large in the respondent-driven sample if it has strong inbreeding and if members of other groups have weak inbreeding. Conversely, a large group might appear small in the respondent-driven sample if its inbreeding is weak, and if other groups have strong inbreeding. The effectiveness of the RDS as a means for drawing unbiased samples depends on whether these two possibilities are plausible.

The conditions under which RDS produces unbiased samples can be deduced through analysis of equation 11 above. This can be done by identifying the conditions under which E_a equals P_a . First, substitute P_a for E_a in equation 11 above:

$$P_a = \frac{P_a(I_b - 1)}{P_a I_b - P_a I_a + I_a - 1} \quad (12)$$

then, solve for I_a ,

$$I_a = \frac{I_b(P_a - 1)}{P_a - 1} \quad (13)$$

which can be simplified as follows:

$$I_a = I_b \quad (14)$$

When described in terms of Figure 4, this conclusion means that when inbreeding is equal, the *inflationary* effects of each group's inbreeding is exactly offset by the *deflationary* effects of other groups' inbreeding. Thus the two inbreeding terms cancel one another, leaving sample size determined exclusively by population size. Though this proof applies to the two-group case it extends to the three-group case, and has been confirmed by simulations in larger systems. This theorem may be formally stated as follows:

THEOREM THREE: A respondent-driven sample draws an unbiased sample if all groups' inbreeding terms are equal, i.e., for any group X, $E_x = P_x$, iff $I_x = I_y$ for any other group Y.

A limitation of the theorem is noteworthy. When inbreeding terms are very large, reflecting mutual isolation of the system's groups, the approach to equilibrium slows, e.g., when the terms exceed .99, scores of recruitment waves may be required to approximate equilibrium. Therefore, given the practical limitations that constrain the number of recruitment waves, equilibrium will be reached only when inbreeding is not extreme. The implication is that when the boundaries separating groups are virtually impassible, RDS should be used to draw samples from *within* such groups, and not *across* them, even should inbreeding terms prove to be equal.

Equality of inbreeding terms is a quite stringent condition. Hence, the question of how inbreeding terms are related to one another is crucial to an assessment of RDS's ability to avoid bias. Unfortunately, biased network theory provides no guidance regarding the relationships among inbreeding terms. However, other sociological theories provide reasons to expect that inbreeding will be at least positively related. A venerable principle from Simmel (1955), is that common enemies enhance group solidarity, implying that Balkanization has an epidemic quality, because when any single group strengthens its boundaries, that induces other groups to do the same. More recent research shows that anything that creates a sense of common fate is sufficient to strengthen group boundaries (Rabbie and Horowitz 1969). An example of this occurs when opportunities for association are determined by group identity. Thus, in a system where other groups strengthen their degree of inbreeding, a group will

The numerator of this expression is positive, because P_a is positive by assumption, $(1 - I_b)$ and $(1 - P_a)$ are positive because $I_b < 1$, and $P_b < 1$, and the product of three positive numbers are positive. Similarly, the denominator is positive because the square of any number is positive. Thus, dE_a/dI_a is positive because its numerator and denominator are both

react by increasing its own inbreeding. Conversely, in a system where inbreeding is weak, the group will tend to avoid inbreeding. The cumulative effect of these processes creates a positive relation among inbreeding terms. If this line of theorizing is correct, the equal-inbreeding assumption is approximated in many social systems. Of course, this speculation requires empirical confirmation.

It might seem that the best way of empirically assessing the bias of a RDS would be to compare it with the population from which it was drawn. Unfortunately, in the current research, this is not possible given the hidden nature of that population. One approach might be to apply the method to a non-hidden population with known characteristics. Such a test would be valuable, but less than definitive. Problems of masking do not arise in non-hidden populations, so the generalizability of the results to hidden populations would remain unclear. A second approach applicable to some hidden populations might be to use very aggressive measures to produce a saturation sample, so the total sample could then be compared with partial samples previously derived from a standard RDS. This more aggressive recruitment mechanism could involve an additional incentive, in which recruiters are rewarded not only for their own activities, but also for those of their recruits. This would create an incentive for recruiters and their recruits to cooperate by pooling their networks and social influence. One indication that such an approach might work is an innovation that spontaneously emerged at both sites. The male dominated drug injection scenes created difficulties for some female recruiters, and without any prompting from staff, the same solution emerged at both sites — some women recruited in pairs. The results were highly effective, and suggest that the robustness of recruitment could be increased if all subjects have incentives to recruit in pairs. This shows the capacity of an RDS to harness and constructively channel drug injectors' creativity and energy.

Table III • Recruitment by Ethnicity in Towns 1, 2, and 3, and Comparisons with Population Distributions*

Comparison of Race/Ethnicity in Sample of Recruits (S), Population Living Below Poverty Line (P), and Total Population (T) by Town									
Race/Ethnicity (n)	Town 1			Town 2			Town 3		
	S	P**	T	S	P**	T	S	P**	T
Non-Hispanic Black	27.6%	23.7%	16%	11.4%	27.6%	10.9%	12.1%	9%	3.6%
Non-Hispanic White	54.3%	45.2%	69%	55.7%	61.3%	84.2%	56.9%	52%	82.8%
Hispanic	14.7%	26.7%	11.9%	31.4%	9.6%	3%	27.6%	39%	12.9%
Other	3.4%	4.4%	3.1%	1.4%	1.5%	1.9%	3.4%	0%	.7%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%
	(93)	(4,310)	(28,540)	(70)	(2,993)	(42,762)	(58)	(4,342)	(59,479)
Mean Discrepancy (D)	Mean Discrepancy Between Recruits and: Poor Population, D = 6.5% (r = .926) Total Population D = 7.4% (r = .957)			Mean Discrepancy Between Recruits and: Poor Population, D = 10.92% (r = .804) Total Population, D = 14.5% (r = .946)			Mean Discrepancy Between Recruits and: Poor Population, D = 5.7% (r = .950) Total Population D = 13.0% (r = .943)		

Notes:

* Source, 1990 Census

** Definition of poverty - family of four earning below \$12,674 for 1989

Available data permit only rough comparisons between the sample distributions and the hidden populations from which they are drawn. This involves comparing samples of recruits

to the demographics of their communities, on the presumption that the hidden population will to some extent mirror the larger community. Table III shows comparisons of racial and ethnic composition between the sample of recruits and the racial and ethnic distribution of the total population, and between the sample and the distribution of the population living below the poverty line. The latter comparison is potentially more relevant, because injectors come disproportionately from the ranks of the poor. For example, in aggregate at both sites, more than three quarters of subjects reported sub-poverty line levels of income. This analysis focuses only on subjects living within the sites' three principal towns. Subjects scattered in surrounding and more distant communities were excluded, because too few subjects came from any single community to provide a meaningful sample size.

Though it is important not to make too much of these tentative comparisons, several findings are notable from inspection of Table III. First, as measured by mean discrepancies between the sample and population, each sample corresponds more closely to the distribution of persons living below the poverty line than to the total population. This no doubt reflects differential recruitment of the poor into drug injection. Second, a major discrepancy exists in Town 2 between the proportion of Hispanics in the sample (31.4%) and in both the total population (3%), and the poverty population (9.6%). This may constitute a substantial over-sampling of Hispanics of from 10.5 to 3.2 times. This discrepancy may result from any number of sources: changes in the Town's ethnic composition since the 1990 census, in which case sampling is correct; special attributes of the town's Hispanic population that introduce some unknown form of bias into the sample; or an artifact of incomplete sampling. In any case, this form of sampling does not appear to undersample minorities. Respondent-driven sampling continues at this site, and ethnographic field investigation has begun, so these further investigations may resolve the issue.

Finally, in Towns 1 and 3 the comparison between the sample and poor populations is relatively close, i.e., mean discrepancies equal only 6.5% and 5.7% respectively. Even in Town 2, where the apparent over-sampling of Hispanics constitutes an anomaly, the comparison becomes relatively close if one excludes Hispanics. Therefore, in these cases RDS appears to succeed in producing samples that include a broad cross section of the underlying population, and hence is "representative" in Spreen's and Zwaagstra's sense.

Sensitivity Analysis

Given the importance of the equal inbreeding condition for RDS's ability to draw unbiased samples, and given that however strongly related the inbreeding terms may prove to be, they can never be expected to coincide exactly, it is useful to perform a sensitivity analysis to determine what happens when the equal inbreeding requirement is violated. Figure 4a reports the results of simulations in systems composed of four groups.⁵ The vertical axis represents the mean absolute discrepancy between population and equilibrium sample distributions. This term can vary from zero, when the two distributions are identical, to a

5. The simulation experiments reported in Figure 4a involved a seven step procedure:

- (1) The population distribution is set randomly. This was done by choosing four random numbers between 0 and 1, and dividing each by the sum of the four numbers to yield four positive numbers that sum to one. These numbers represent the size of each of the four groups, i.e. P_a , P_b , P_c and P_d .
- (2) Each point on Figure 6a's horizontal axis corresponds to a *range* within which inbreeding must fall. The *midpoint* of this range is chosen randomly, within the set of values consistent with the requirement that inbreeding cannot exceed 1, or be less than 0. For example, if the range chosen is .6, the midpoint of the interval must lie between points .3 and .7.
- (3) Inbreeding terms (I_a , I_b , I_c , and I_d) are selected randomly from within the above-chosen range, e.g., if the midpoint of the range is .4, and the length is .6, four random numbers are chosen between .1 and .7.
- (4) Selection probabilities are computed using equations 5 and 6.

maximum value of .5 in a system with four groups (i.e., for a system of $n = 4$ groups, the maximum is $2/n = 2/4 = .5$). The horizontal axis represents the range within which inbreeding terms fall. This range's lower limit is zero, indicating that all inbreeding terms are equal. The range's upper limit is 1, indicating that inbreeding may vary within the full theoretically allowable range of 0 to 1, and hence are maximally diverse. To reflect some of the resource limitations that necessarily attend actual research, recruitment in the simulations began with a homogeneous set of initial subjects (the worst-case possibility), and continued for 10 waves. Each simulation was repeated 1,000 times to determine the mean absolute discrepancy between sample and population (see the bold line in Figure 4a). The area bounded by the dotted lines represents the confidence interval for the discrepancy. When the range is zero, and hence the inbreeding terms are all equal, discrepancies are minimized (i.e., $3.3\% \pm 1.5\%$). There is not a perfect association between sample and population because of the limitation in the number of waves. It is also apparent by inspection of Figure 4a that as the differences among inbreeding grows, sampling discrepancy increases. However, even when inbreeding is maximally diverse, with a range equal to one, a positive association between sample and population continues to exist, as reflected in a sampling discrepancy (i.e., $12.8\% \pm 7.7\%$) that is substantially less than the theoretically possible maximum. This corresponds to a correlation between sample and population distributions of $r = .57 \pm .114$. Thus, the results show that, even in the worst-case scenario where inbreeding is most disparate, a RDS nonetheless produces a sample that contains a cross-section of the underlying population.

A more complete characterization of the effects of violations of the equal inbreeding assumption concerns the effect of the *magnitude* of the terms. From theorem three we know that if inbreeding is equal and the number of waves is sufficient to reach equilibrium, sampling is unbiased. However, a question remains: when inbreeding is unequal, does the strength of inbreeding affect sampling? Figure 4b depicts the results of a series of simulations that answer this question. As before, the vertical axis represents the discrepancy between the population and the sample. However, the horizontal axis now represents the magnitude of inbreeding. This is operationalized by first specifying the range within which the inbreeding terms fall. This was set at a range of .5, the mean theoretically possible value. Inbreeding terms within this range were then allowed to vary from a minimum midpoint value of .25 (a smaller midpoint value would not be possible because then some inbreeding would be negative), and a maximum midpoint value of .75 (a larger midpoint value would not be possible because then some inbreeding would exceed 1). As before, the bold line represents the discrepancy between sample and population, and the dotted line represents the confidence interval. RDS yields the most biased samples when inbreeding is greatest. However, even in this worst case scenario, the association between sample and population remains positive and substantial, as reflected in a mean discrepancy of $12.1\% \pm 8\%$. This corresponds to a correlation between sample and population distributions of $r = .55$. When inbreeding is weaker (e.g., less than $1/3$), RDS yields substantially less biased samples (i.e., mean discrepancies are less than 4%).

In sum, the sensitivity analysis shows that departures from the equal inbreeding assumption must be substantial before they introduce substantial biases into the sample. The form of departure that has the most adverse effect occurs when inbreeding is very strong, and yet even under these conditions the populations and sample distributions remain associated.

(5) Equilibrium sample distributions were estimated using the procedure described for Figure 3. To reflect the resource limitations that exist in most studies, recruitment was assumed to continue through only ten waves, and the initial seeds were assumed to come from the same group.

(6) The mean absolute difference between the estimated equilibrium distribution (i.e., the composition of the 10th wave of recruits), and the population distribution was computed.

(7) Steps 1 to 6 above were repeated 1,000 times, then the mean and standard deviation of the mean differences were computed.

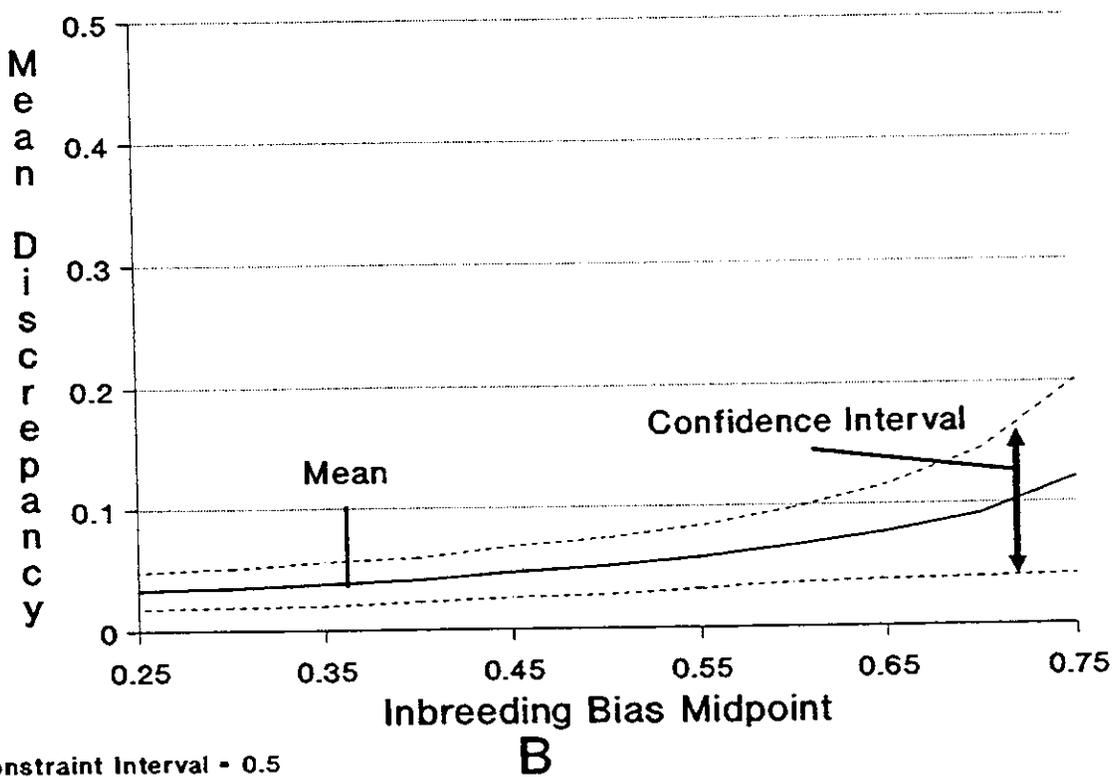
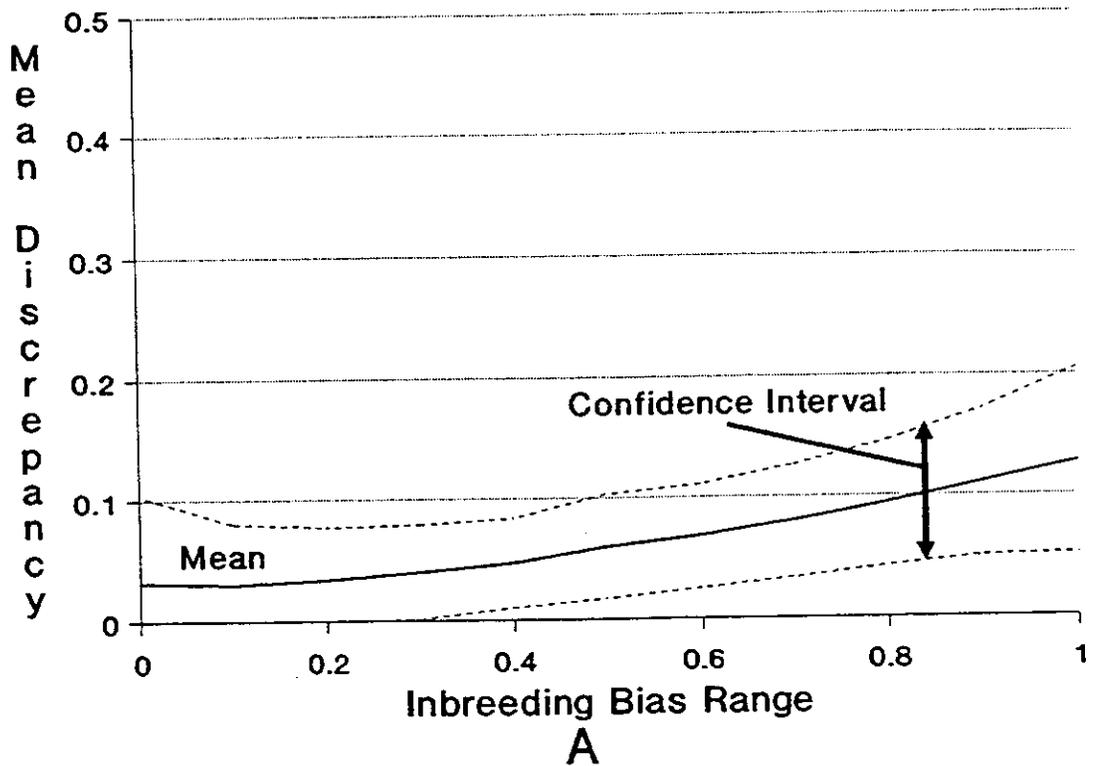


Figure 4 • Population/Sample Discrepancy as a Function of Constraints on Inbreeding Biases

Hence, even when the equal-inbreeding assumption is violated, RDS can be expected to produce good cross-sections of the target population.

Limitations of the above analysis should be noted. Recall Spreen and Zwaagstra's argument that samples of hidden populations cannot expect to do more than offer broad cross-sections of the target population. The above analysis suggests that RDS fulfills this goal, and that more may be possible. As shown above, biases in the first step of recruitment tend to decrease as the successive waves are recruited. Furthermore, under a broad array of circumstances (i.e., less than extreme and positively associated inbreeding levels), RDS can be theoretically expected to produce samples with only a modest degree of bias. Potentially, if means could be found for independently assessing inbreeding levels, samples could be weighted to eliminate these biases. In addition, steering incentives, which provide a means for altering inbreeding levels, could be used to either over-sample groups of special interest, or to make inbreeding levels more nearly equal and thereby reduce the need to weight the samples. Another major remaining task is the investigation of these samples' consistency and the resulting variability of estimators. Only then will it be possible to compute standard errors and confidence intervals for population estimates, and thereby make RDS into a fully defensible statistical sampling procedure.

Conclusion

This paper presented empirical results from a new approach to accessing and sampling hidden populations, designed to reduce several deficiencies afflicting traditional forms of chain-referral samples. First, it is usually assumed that inferences about individuals must rely mainly on the initial sample, since additional individuals found by tracing chains are never found randomly or even with known biases. However, if the sampling process is allowed to continue through enough waves to reach equilibrium, its composition will be independent of the initial subjects. A second problem is that chain-referral samples tend to be biased toward more cooperative subjects. This problem can be reduced by RDS's dual incentive system. Individuals who resist researchers' appeals may nonetheless yield to appeals from their peers. A third, related problem is that chain-referral samples are biased by "masking" (protecting friends by not referring them). This problem can be reduced because recruitment incentives weaken the reluctance to approach reclusive peers. Finally, chain-referral samples may be biased toward subjects with large personal networks. There are several ways in which this problem can be addressed: weighting samples based on network sizes, saturating targeted areas, or using steering incentives to increase recruitment of subjects with traits associated with small personal networks.

RDS need not be employed alone. It also can be applied in combination with other methods. Consider first network sampling (Sudman et al 1988). This form of sampling increases the amount of information obtained from each respondent by including questions about their personal networks. Ideally, respondents are selected by a procedure that results in each population member having a known probability of selection, but this is not possible if the population is hidden. Yet the implication of theorems one and two is that the choice of the initial sample need not introduce bias if a suitable procedure is followed. That is, the sampling should first proceed through multiple waves. This permits computation of recruitment networks as depicted in Table I. This, together with knowledge of the composition of the initial set of seeds, makes it possible to compute how many waves must be completed before the sample approximates the equilibrium distribution, and thereby becomes independent of its starting point. Usually this will be no more than three to five waves. Thus, the RDS analysis provides a principled basis for deciding how many waves must be completed before whatever biases were present in the selection of seeds is overcome. The subjects for

the network sample are then drawn from that or subsequent waves. In essence, respondent-driven sampling serves as the initial stage of the sampling procedure, which leads from an initial set of subjects with an unknown bias, to a subsequent set of subjects that are independent of that bias.

RDS may be able to play a similar initial role in ethnographic investigation. When dealing with hidden populations, gaining access to a suitably diverse set of ethnographic informants is frequently a lengthy and uncertain process. RDS may provide a means both for speeding this process, and for ensuring that different sectors of the population are adequately represented among the informants. Thus, as in the case of RDS's combination with network sampling, the number of waves required to reach equilibrium should be computed, and then ethnographic informants are drawn from that and subsequent waves, until the desired number of informants is attained.

References

- Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland
1975 *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass: MIT Press.
- Broadhead, Robert S., and Douglas D. Heckathorn
1994 "AIDS prevention outreach among injection drug users: Agency problems and new approaches." *Social Problems* 41:473-495.
- Broadhead, Robert S., Douglas D. Heckathorn, Jean-Paul C. Grund, L. Synn Stern, and Denise L. Anthony
1995 "Drug users versus outreach workers in combating AIDS: Preliminary results of a peer-driven intervention." *Journal of Drug Issues* 25(3) 531-564.
- Coser, Lewis A
1956 *The Functions of Social Conflict*. Glencoe, Ill.: Free Press.
- Deaux, Edward, and John W. Callaghan
1985 "Key informant versus self-report estimates of health behavior." *Evaluation Review* 9:365-368.
- Erickson, Bonnie H
1979 "Some problems of inference from chain data." *Sociological Methodology* 10:276-302.
- Fararo, Thomas J
1973 *Mathematical Sociology: An Introduction to Fundamentals*. New York: John Wiley and Sons.
- Fararo, Thomas J., and John Skvoretz
1984 "Biased networks and social structure theorems: Part II." *Social Networks* 6:223-258.
- Frank, Ove, and Tom Snijders
1994 "Estimating the size of hidden populations using snowball sampling." *Journal of Official Statistics* 10:53-67.
- Goodman, L. A.
1961 "Snowball Sampling." *Annals of Mathematical Statistics* 32:148-170.
- Heckathorn, Douglas D.
1990 "Collective sanctions and compliance norms: A formal theory of group-mediated social control." *American Sociological Review* 55:366-384.
1993 "Collective action and group heterogeneity: Voluntary provision versus selective incentives." *American Sociological Review* 58:329-350.
1996 "Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61:250-278.
- Heckathorn, Douglas D., and Robert S. Broadhead.
1996 "Rational choice, public policy, and AIDS." *Rationality and Society* 8:235-260.
- Kemeny, John G., and J. Laurie Snell
1960 *Finite Markov Chains*. Princeton, N.J.: Van Nostrand.

- Killworth, Peter D., and H. Russell Bernard
1978/79 "The reversal small world experiment." *Social Networks* 1:159-192.
- Klov Dahl, Alden. S.
1989 "Urban social networks: Some methodological problems and possibilities." In *The Small World*, M. Kochen (ed.), 176-210. Norwood, N.J.: Ablex.
- Rabbie, Jacob M., and Murray Horwitz
1969 "Arousal of in-group-out-group bias by a chance win or loss." *Journal of Personality and Social Psychology* 13:269-277.
- Rapoport, Anatol
1979 "A probabilistic approach to networks." *Social Networks* 2:1-18.
- Simmel, Georg
1955 *Conflict: The Web of Group Affiliations*. Trans. Kurt H. Wolff and Reinhard Bendix. New York: Free Press.
- Spreen, Marius
1992 "Rare populations, hidden populations, and link-tracing designs: What and why?" *Bulletin de Methodologie Sociologique* 6:34-58.
- Spreen, Marius, and Ronald Zwaagstra
1994 "Personal network sampling, outdegree analysis and multilevel analysis: Introducing the network concept in studies of hidden populations." *International Sociology* 9:475-491.
- Sudman, Seymour, and Graham Kalton
1986 "New developments in the sampling of special populations." *Annual Review of Sociology* 12:401-429.
- Watters, John K., and Patrick Biernacki
1989 "Targeted sampling: Options for the study of hidden populations." *Social Problems* 36(4):416-430.
- Welch, Susan
1975 "Sampling by referral in a dispersed population." *Public Opinion Quarterly* 39:237-245.