Monitoring AIDS and Other Rare Population Events: A Network Approach[1]

Edward O. Laumann
The University of Chicago and NORC

John H. Gagnon
State University of New York and NORC

Stuart Michaels
The University of Chicago and NORC

Robert T. Michael
The University of Chicago and NORC

L. Philip Schumm
The University of Chicago

June, 1991

[1]Revision of a paper presented at the Sunbelt Conference of the International Network for Social Network Analysis, Tampa, Florida, February 14-17, 1991. Some earlier results were presented at the Annual Meeting of the American Association for the Advancement of Science, New Orleans, Louisiana, February 15-20, 1990. The GSS is funded by the National Science Foundation; the 1988, 1989 and 1990 questions pertaining to sexual behavior and acquaintance with AIDS, homicide and suicide victims were funded by the National Opinion Research Center (NORC). Partial support for this research was provided by the National Science Foundation through the National Data Program for the Social Sciences, NORC, and the National Institute of Child Health and Human Development (contract N01-HD-8-2907). We are indebted to Patricia Styer for her technical contributions to earlier drafts of the section on trends.

Abstract

The U.S. Public Health Service systems responsible for monitoring the occurrence of health events are, of necessity, decentralized and highly bureaucratic in nature. Consequently, it is useful to ask how accurate these systems are in providing us with information about the incidence of health events and about the composition of those who experience these events. In this paper, we propose a new technique for devising independent estimates of the relative prevalence of health events across population subgroups. This technique uses information provided by a random sample of respondents about the occurrence of a particular health event within their acquaintance networks to infer the demographic composition of all similarly affected persons in the population. Data from the 1988, 1989, and 1990 General Social Surveys (GSS) on respondents' acquaintance with homicide victims, suicide victims, and persons with AIDS are compared to the official statistics published for these three events. While the GSS data reproduce fairly closely the official homicide and suicide statistics for sex and region, there are significant discrepancies with respect to age and race. In all three years, the distributions of persons with AIDS in GSS respondents' networks show a higher proportion of AIDS cases among whites, a higher proportion in the Midwest, and a lower proportion in the East than the official statistics. Characteristics of both the official reporting apparatus and our network method that might account for these discrepancies are discussed. In addition, time series data from the National Health Interview Survey measuring the proportion of persons knowing someone with AIDS are used to explore patterns in the spread of the disease over time.

Introduction

In the United States, official systems responsible for monitoring the occurrence of certain health related events are often highly decentralized. Rather than gathering data directly from the affected persons, government agencies such as the Centers for Disease Control (CDC), the National Center for Health Statistics (NCHS), and the Federal Bureau of Investigation (FBI) must rely on loosely coupled, multi-level, multi-agency reporting systems situated in a variety of jurisdictions. The nature of these systems makes them vulnerable to systematic distortions and overt manipulations by interested parties at various levels of the reporting hierarchy (Shilts, 1987; Laumann and Knoke, 1987).

A good example of these decentralized monitoring systems is the surveillance system set in place by the CDC to scan the society in order to identify health events that qualify as AIDS cases (Curran et al., 1988). This surveillance system is composed of three stages: the case gathering process, the case reporting process, and the case classification process (GAO, 1989).

The case gathering process has two steps: first, the construction of a clinical definition of AIDS, and second, the application of this definition by various medical professionals when making diagnoses. In order to distinguish between AIDS and other diseases and also between full blown AIDS and earlier stages of HIV disease, the CDC has identified a list of conditions, symptoms, and diagnostic tests that are required in order for a health event to be diagnosed as AIDS. However, this definition has not remained stable over the course of the epidemic for both medical and non-medical reasons. For example, in 1987 the original case definition was revised to include "wasting syndrome" and dementia as conditions that qualify for an AIDS diagnosis and there is pressure at the present to include additional conditions, particularly those associated with HIV disease in women. Much of this pressure on the CDC to expand the set of conditions necessary for an AIDS diagnosis results from the fact that having AIDS qualifies a person to receive health and welfare benefits such as disability, Medicaid, Social Security income, and welfare. In addition to changes in the official definition of AIDS, the CDC also accepts presumptive diagnoses (without laboratory tests) of the necessary conditions as legitimate AIDS cases, as long as the patient has also tested positive for HIV. This may allow physicians to make the diagnosis more quickly in situations where the patient can benefit from being diagnosed.

The activities of individual physicians, coroners, and other health professionals comprise the second step in the case gathering process. At this point non-reporting may occur because of the normal errors of misdiagnosis, excessive and lost paperwork, the failure of infected persons to seek treatment, and minimum incentives to diagnose certain cases (e.g., a patient dies so quickly that there is no clinical reason to make the diagnosis). Another problem is that the social stigma attached to AIDS and to those most often infected by the disease may act as an incentive for physicians not to make the diagnosis. This process would operate to conceal the disease among those social groups whose members would sustain the greatest reputational losses if diagnosed and who are able to pay for private health care. While AIDS is also stigmatizing to those with the fewest economic and political resources, these groups are likely to come into contact with the routine reporting systems (e.g., hospitals), have the least ability to manipulate the reporting systems, and have an incentive to be diagnosed in order to gain access to contingent health and welfare benefits.

During the case reporting process, local health authorities are responsible for collecting information about specific cases from the diagnosing agents (physicians, coroners, and hospitals). These local authorities may or may not impose active surveillance and systematic record checking on the diagnosing agents in their area. Such vigilance on the part of local authorities will determine when and if cases are reported, the quantity and accuracy of additional information reported about each case (social demographic characteristics and information about risk group), and the speed with which reported cases are forwarded to the CDC (Conway et al., 1989; Modesitt et al., 1990). Although we would expect to find highly vigilant systems in areas with large numbers of AIDS cases, there is some evidence that the systems in these areas suffer fatigue as the numbers of cases grow very large and the disease becomes more routinely identified. Moreover, there may be incentives on the part of hospitals and other health care facilities either to report or not to report

AIDS cases. On the one hand, if drugs for treatment are only available in research trials and hospitals' reimbursements are higher for AIDS cases, the likelihood that the disease will be reported may increase. Other incentives for reporting may be created as new Federal programs allocate additional funds to the hardest hit cities under the "Ryan White Comprehensive AIDS Resources Emergency Act of 1990." On the other hand, if hospitals are concerned about a loss of business due to the public's fear of AIDS, there will be a countervailing pressure not to report the disease.

As a final step in the surveillance system, the CDC examines each case that is reported to them and classifies it according to a set of criteria developed to estimate the original method of infection (membership in a "risk group"). The accuracy of these estimates depends upon the thoroughness and accuracy of the information reported about each case, and upon the CDC's ability to discriminate between various possible methods of infection.

As we have outlined above, the decentralization of the CDC's AIDS monitoring system creates opportunities at each level for interested parties to attempt to influence the system. In fact, concerns about the accuracy and timeliness of the epidemiological estimates of the number of AIDS cases have been expressed since the opening stages of the epidemic (Institute of Medicine, 1986; Turner et al., 1989). Some of the original sources of these concerns, e.g., the novelty of the disease, the uncertainty about its modes of transmission, and the actual agent of transmission, have now been eliminated. However, problems associated with the incentives created by funding policies, the stigma attached to those who have been infected, and the marginalized social status of those who are most often infected have not yet been solved. These problems may produce not only an imprecise estimate of the total number of AIDS cases, but also an inaccurate picture of the distribution of AIDS cases across exposure groups and various socio-demographic groups.

We refer to this distribution of AIDS cases as the social epidemiology of the disease. More precisely, we define social epidemiology to be the distribution in social space of a particular health event, which is acquired by individuals and is facilitated or hindered by the operation of social processes.[1] Increasing the accuracy with which we measure the social epidemiology of a health event such as AIDS is important for understanding the disease, projecting its future, hindering its course, and mobilizing resources. For instance, the accuracy of the data on AIDS cases is particularly important since they are a primary source of estimates about the future impact of the disease. Changes in the number of AIDS cases, their exposure, and background characteristics are the primary measurements for estimating changes in the composition of those infected even though the AIDS cases that are tabulated in a given year may be representative of infections contracted some years before.[2] Estimates of those infected do not entirely depend on these estimates of the numbers of AIDS cases, but the weaknesses of the data-gathering system for asymptomatic HIV infection are so great that estimates of the numbers of cases with frank AIDS remains crucial to estimating the size and composition of the epidemic.

The CDC has attempted to strengthen its AIDS monitoring system through both economic and educational support for reporting agencies and individuals and specialized quality control investigations, yet these measures are not tantamount to effective control. To deal with this issue, the CDC is now engaging in monitoring the AIDS surveillance system as well as de-emphasizing the singular importance of the case reporting system by attempts to develop a surveillance system that will more directly monitor levels of HIV infection through the "Family of Surveys" (Dondero et al., 1988). These are important attempts at surveillance, however only one of the programs collects systematic data (testing the heelstick blood of newborns to assess the HIV status of the mother) and all other studies rely on samples that are subject to considerable selection bias. In any

[1]The term "social epidemiology" is often used by epidemiologists to denote the study of the role of social factors in disease (Feinstein, 1985). Although we have restricted our definition to the actual distribution of an event in social space, examination of this distribution can help to identify the social processes that channel the event's occurrence. Therefore, we view studying the social epidemiology of an event as a necessary antecedent to more focused research aimed at exploring those processes.

[2]The latency period from infection to the appearance of those symptoms that are the markers of frank AIDS appears at present to be close to ten years, based on a limited number of studies of gay men or hemophiliacs. There is good reason to believe that this latency period may differ among different groups based on health status at infection and quality of life after infection. Thus, the latency period may be sensitive to levels of stress, nutrition, exercise, social support, health care for non-HIV related illnesses, and drug and alcohol use. Moreover, latency periods may have changed for specific exposure groups because of behavioral changes during the epidemic. Two examples may be suggested for gay men: reductions in anal sex may have reduced exposure to different strains of HIV (as well as other sexually transmitted diseases) while changes in fast track lifestyles have reduced the negative effects of alcohol and drug use and limited rest and exercise on the immune system.

case, such surveillance of HIV infection will not replace the need to develop independent methods for monitoring AIDS cases in order to assess the strengths and weaknesses of particular estimates and projections.

Similar problems to those that we have outlined with regard to the CDC's AIDS monitoring system are present in other official reporting systems monitoring the occurrence of various health problems, diseases, and criminal behaviors. For this reason, we believe that it is important to construct independent estimates of the overall prevalence and relative prevalence across population subgroups of these health events. Independent estimates could be used not only to validate official reports, but also to identify areas in which the official reports are particularly weak.

Our Approach

In an attempt to devise independent estimates of the relative prevalence of health events across population subgroups, we adopted a rationale resting on the social network perspective as it is applied to randomly sampled population surveys (Laumann, 1973; Mitchell, 1969; Holland and Leinhardt, 1979; Fischer, 1982; Burt, 1980; Burt and Minor, 1983; Berkowitz, 1982; Wellman and Berkowitz, 1988; Laumann et al., 1989). The strategy is to ask an individual with a known probability of selection from a well-defined population to scan his or her primary acquaintance network, defined to include all the persons he or she knows personally such as kin, friends, neighbors, co-workers, and more casual and incidental acquaintances, in order to identify all those who possess a particular characteristic such as a health condition like AIDS or being the victim of a particular crime. This task that the individual is asked to perform is analogous to that of the CDC case-gathering process; the individual must have criteria for "diagnosing" AIDS and be alert to the necessary signals among his or her social network. For most people the size of such a network is fairly large (on the order of 2,000 to 6,000 persons) (Boissevain, 1974; Pool and Kochen, 1978; Freeman and Thompson, 1989; Bernard et al., 1987; Bernard et al., 1988).

In the 1988, 1989 and 1990 General Social Surveys, respondents were asked questions about their acquaintance with someone, living or dead, who had contracted the disease called AIDS. Inquiries were made about how many such persons the respondent knew. The respondent was then asked additional questions about the person with AIDS he or she knew best, including the nature of the tie (lover, kin, co-worker, or other relationship) and the age, sex, and race or Hispanic origin of that person. In the 1989 and 1990 surveys, the respondent was asked to provide this demographic information on all the individuals whom he or she believed had contracted AIDS, and in addition, he or she was asked the state of residence of each person with AIDS. An identical set of questions (with corresponding changes between the 1988 versus the 1989 and 1990 versions) concerning the respondent's acquaintance with the victim or victims of a (willful) homicide within the last twelve months was also asked. In 1990 a set of questions regarding acquaintance with suicide victims in the past year was added as well.

Our approach is to use the demographic composition of the sets of AIDS cases, homicide victims, and suicide victims reported by the GSS respondents to infer the composition of similarly affected persons in the general population. Discrepancies between our estimates and those reported by the official agencies may reveal inaccuracies in the official reports. Such discrepancies may also reveal differences in the social construction and definition of the particular illness or event. For example, at this time the CDC maintains that the presence of certain vaginal infections often accompanying AIDS is not alone sufficient for an official diagnosis of AIDS. However, if a doctor informs an HIV positive patient that she has an infection (e.g., vaginal candidiasis) that is related to AIDS, she may understand that to mean that she has AIDS. Similarly, an elderly person who is terminally ill and decides to swallow the contents of a full bottle of sleeping pills might be viewed by his or her family as dying of natural causes, while the coroner would diagnose the death as a suicide. These examples illustrate the potential impact of the social definition of an event on the results of both official reporting systems and our network monitoring technique.

The yearly incidence of homicides and suicides are comparable to the prevalence of AIDS in terms of their relative rarity and demographic incidence (for example, in the over-representation among young males). Because homicides and suicides are generally considered accurately reported in the official statistics, if the essential features of the annual homicide or suicide incidence can be reproduced by our network projection techniques from the GSS sample reports, subject to variability in estimates because of the relatively small sample size, then our estimates made with respect to the prevalence of AIDS can be taken more seriously. In addition, all three events have very different social significances attached to them and involve quite different processes of social construction, both within the medical community and within the larger population. Comparing the official statistics and our GSS results for these three events simultaneously will allow us to speculate on the impact of these social significances on the operation of the different monitoring systems.

More precisely, our technique can be likened to drawing a sample of persons from the population corresponding to the total number of acquaintances known to the original GSS respondents. This "network" sample[1] is analogous to a cluster sample, with the acquaintance net of each focal GSS respondent representing a unique cluster. This idea can be illustrated using the question on persons with AIDS (PWAs). Our data consist solely of a count, for each cluster (i.e., GSS respondent), of the number of PWAs in that respondent's acquaintance net. Additionally, we have data describing certain characteristics of those PWAs. Yet we do not know the total size of the network sample, nor do we know the composition of the entire network sample. For statistical purposes, cluster sampling is less efficient than simple random sampling because observations within the same cluster violate the assumption of independence. Although methods for assessing the efficiency of cluster samples have been developed (Altham, 1976; Cohen, 1976; and Brier, 1980), they require knowledge of the whole sample (not just PWAs) that we do not have. Thus, we are unable to determine the statistical significance of differences in the proportion of PWAs in our network sample across population subgroups.

---

[1]It is important to note that the term "network" sample as we have used it here is quite different from the concept of network sampling that is used by network researchers. In that context, network sampling represents a means by which global network structures are inferred from the information given by a sample of respondents.

A similar approach to sampling rare populations called "multiplicity sampling" has been discussed by Sirken (1970, 1972), Sudman (1986), and Kalton and Anderson (1986). In fact, the theory behind multiplicity sampling has been developed well enough to make unbiased estimates of sample statistics and to compute sampling variances. Multiplicity samples are drawn using either respondents' kinship networks or other explicitly delineated networks linking the respondent to a group of others. This allows a researcher to collect information about the size of the network that each respondent is reporting on (referred to as that respondent's "multiplicity.") These sizes can then be used to estimate the sampling variation in the total number of persons reported by the respondents to have been affected by a particular event. In our case, the multiplicity rule used by the GSS (all acquaintances of the respondent) specified personal networks that were too large for the respondent to enumerate.[1]

In addition to the statistical problems inherent in the GSS data, we are also aware of several potential biases in the network monitoring procedure that may affect our conclusions. An important source of bias arises from the fact that personal networks of individuals are known to vary greatly in size, social composition, levels of intimacy and social access, and density. Moreover, these features of social networks may vary systematically across social groupings in the society -- the networks of women may vary from those of men, members of different ethnic groups may vary from each other (in the case of AIDS African Americans may differ from Euro-Americans, Hispanic Americans, and others), and members of different risk groups may have different social networks.

If, for example, people draw almost all of their acquaintances from their own racial groups, systematic variation in the size of acquaintance nets across those racial groups could produce biased estimates of the relative prevalence of AIDS cases with respect to race. Likewise, systematic differences between racial groups in the levels of intimacy and access to health status information might alter both the quantity and type of information that a respondent has about his or

[1]Sirken (1970) suggests using a multiplicity rule based either on consanguine relationships or on spatial proximity. However it is unlikely that these multiplicity rules, used with a sample of 2,000 or less respondents, would generate enough AIDS cases or homicide cases to analyze.

her acquaintances. Differences in levels and kinds of personal information about network members and systematic differences in knowledge about HIV/AIDS might in turn produce systematic differences between respondents in their ability to "diagnose" AIDS within their acquaintance nets accurately. These factors might produce biased estimates of the relative prevalence of AIDS within the specific population subgroups in question. However, as a first approximation, we shall make the assumption that, on the average, these differences in network structure are not systematically organized by the social characteristics that are of special interest to us. Our ability to reproduce much of the official profile for homicide and suicide suggest that this first approximation is reasonable.

Yet another more complex issue is whether or not a comparison between our network procedure and cluster sampling can accurately be made. Traugott, Groves, and Downes-Le Guin (1990) have made the point that slight variations in the wording of the question "How many people do you know with AIDS?" can produce differences in the number of cases reported.[1] One way to explain these findings is that peoples' acquaintance networks are imprecisely bounded at the margins, and that question wording is important in determining how respondents define their acquaintance networks. This point also raises a more subtle issue; namely, what is the cognitive process by which respondents answer this question? Do they, according to our cluster sample analogy, scan through their list of acquaintances, objectively "diagnosing" each person for AIDS? Or, do they instead think of whole groups of individuals at a time, scrutinizing only a few potential PWAs?

Obviously the time and memory constraints are too great for any respondent to scan through their list of acquaintances like a computer scanning data. However, based on three assumptions we believe that the cluster sample analogy is still reasonable. First, we assume that

---

[1]These researchers report differences of up to 14 percentage points in the percentages of respondents claiming to know someone with AIDS across nine national surveys by different survey organizations that employed different question wordings. Unfortunately, they do not discuss the obvious fact that these surveys were conducted over a period of five years (1983 to 1988) when AIDS cases known to the CDC were doubling at almost yearly intervals. One would expect an upward secular trend over this period of time, and, in fact, one observes it (setting aside the outlying ABC survey of September 1985 that claims 16 percent of the population knew someone with AIDS - a percentage almost 3 times higher than any other survey near it in time.)

most people only report PWAs with whom they have had some social contact (direct or indirect.) This means that respondents are not reporting public figures with AIDS whom they do not know personally.[1] Second, respondents "diagnose" an acquaintance as having AIDS according to some medical indicator (HIV positive, a physician's diagnosis, etc.), and do not report systematically organized false positives (e.g., a middle aged, never married man who looks very thin and ill). Third, although some PWAs might conceal information about their diagnoses from acquaintances and respondents might conceal acquaintances with AIDS in survey interviews, we assume again that there are similar rates of these behaviors across the population subgroups that are of interest to us.

This brief treatment of the potential biases inherent in our network monitoring technique is not intended to prove that conclusions about the relative prevalence of AIDS in population subgroups based on this technique are completely accurate representations of "reality" or should substitute for the CDC representation. Rather, our intention is to convince the reader that large differences between the official statistics and the alternative figures we present merit an explanation. In addition, although we believe that the current GSS AIDS items do not generate sufficient data to support refinements to the simple technique we present here, we hope that our discussion of possible problems will stimulate additional work on the use of network methods for monitoring purposes.

Relative Prevalence Across Subgroups: GSS versus Official Statistics

The GSS is a national area probability sample of about 1,500 households conducted by the National Opinion Research Center (NORC); the face-to-face survey has been conducted annually

[1]This seems reasonable given the relatively smooth profile of the plot of the percentages who know a PWA against time (monthly intervals). If significant numbers of respondents were claiming to know public figures with AIDS, we would expect to see upward "steps" in the plot corresponding to the publicity surrounding public figures' diagnoses.

nearly every year since 1972 and is widely used in research in the social sciences. Respondents are randomly selected adults age 18 and over, one from each household. We use the GSS data from 1988, 1989, and 1990; the surveys were conducted between February and April each year. The sample totals for these three years were 1,481, 1,537, and 1,372 completed cases, yielding response rates of 77.3%, 77.6%, and 73.9%, respectively, well within the usual range of response rates obtained for the 17 annual surveys. The GSS compare quite closely with decennial census data and Current Population Survey data on the demographic and economic characteristics of the U.S. population (Davis and Smith, 1989; Bureau of Census, 1983; Current Population Reports, 1988).

Table 1 presents homicide data from official statistical sources and GSS 1988, 1989, and 1990.

Insert Table 1 about here.

The most recent publications of homicide data in the Uniform Crime Report (UCR) (U.S. Department of Justice, 1990) and the Vital Statistics of the United States (National Center for Health Statistics, 1989) are for 1989 and 1987, respectively; this means there is a 1 to 3 year discrepancy between the reference year and the three GSS reports. However, the percentages we are interested in comparing are remarkably stable over time, as an inspection of the annual demographic breakdowns since 1981 will readily demonstrate. Despite small substantive differences in the constituent organizations doing the reporting (police departments versus coroner offices), the two sets of official data reveal very similar profiles of homicide victims.[1]

About 10% of the 1988, 1989, and 1990 GSS samples claimed to know one or more persons who were victims of homicide within the last 12 months, resulting in 255, 238, and 197 characterizations of victims, respectively. These respondents' descriptions of the victims were

---

[1]There is an extensive literature in which the validity and accuracy of the official statistics on homicide and other criminal behavior are debated; however, most researchers agree that homicide is the most reliably reported of the Federal Bureau of Investigation's index of seven serious crimes. (See Sutherland and Cressy, 1978; Nettler, 1978; Gove, et al., 1985.)

used to estimate the relative incidence of homicide by sex, race, age, and geographic location.[1] In the case of sex and region, the GSS estimates approximate the official statistics well: disproportionately more males, more residents of the South, and fewer residents of the East (in particular, New England) are victims of homicide. With respect to age, the GSS samples report slightly more youthful victims than reported by the official statistics, however the official statistics report 20 year old victims in the next higher age category. This difference as well as the well known tendency to round off age reports to the nearest fifth or tenth year probably accounts for the discrepancy between the official and GSS estimates. It is only with regard to race and ethnicity that there appear to be a substantial discrepancy between the official statistics and the GSS estimates. The 1988 and 1989 GSS data suggest a substantially higher percentage of homicide victims are white and a correspondingly much lower proportion are black. The 1990 GSS, however, approximates the official statistics in the racial composition of victims. All three GSS data sets indicate a much lower proportion Hispanic than does the UCR for 1988. These discrepancies are mirrored in the AIDS estimates (see below).

In the case of suicide victimization, for which we have data only from GSS 1990, Table 2 provides the demographic distributions from the Vital Statistics in 1986 and 1987, the last two years for which data are available, together with the estimated distributions from GSS 1990.[2]

---

[1]The GSS asked respondents how many people they knew personally who were victims of homicide in the past 12 months. In 1988, 154 respondents reported knowing one or more victims of homicide, yielding a total of 255 victims. However, respondents were then asked further questions only about the victim they knew best. Therefore, in order to estimate the distribution of the characteristics of homicide victims, we assigned additional victims known to a single respondent the race/ethnicity of that respondent. The distributions of sex and age of victims are based on only those victims described by respondents, since sex and age are probably not the same for both respondent and victim. Region of the victim was not asked, and so it was assumed to be the same as that of the respondent. In 1989 and 1990, respondents were asked to describe all victims of homicide including their region, and these reports were used to tabulate the 238 and 197 victims described. Seven homicide victims (known to one respondent) who died abroad and two who were patients are excluded from the analysis in 1989; three victims were excluded in 1990 because they were patients of the respondents who reported them. We made the decision to exclude patients from our analyses because health professionals are likely to know disproportionately large numbers of affected persons relative to other respondents. The result is that the cases reported by only a few health professionals may greatly influence our estimates of the relative prevalence of health events across population subgroups.

[2]Again, from an inspection of annual data from 1982 through 1987 (not shown), we find remarkably stable demographic distributions from year to year, with variations by categories across years being no more than a percentage point or so in these official reports.

Insert Table 2 about here.

While the distributions for sex and region are quite close, there are substantial differences between the two with respect to race and age. Part of the discrepancy with respect to race is due to the absence of the "Hispanic" category in the Vital Statistics that likely inflates the white percentage. More interesting is the discrepancy in the estimated percentages of black suicides. The GSS suggests the percentage of black suicide victims to be close to the percentage of blacks among the national population, a figure twice as large as the percentage consistently reported in the Vital Statistics since 1982. The notion implied by the Vital Statistics data that blacks are disproportionately less likely to commit suicide strikes us as implausible in light of the severe social, economic, psychological and health disadvantages to which this population is subjected in the United States (Wilson, 1987). The lower official suicide rate among blacks may be a function of discounting certain black deaths that are deliberate as accidents (e.g., drug overdoses) as well as the blurring of suicide into the background of social violence that characterizes the lives of the inner city poor. High death rates from AIDS, tuberculosis, homicide, drug and alcohol use, and accidents among young blacks of both genders may obscure the incidence of suicide and create a callousness in the reporting sensibilities of agencies that then do not carefully differentiate among different causes of death.

More challenging is the observed discrepancy between the Vital Statistics and the GSS with respect to age: GSS respondents report a much higher incidence of teenage suicide than is reported in the Vital Statistics and a much lower rate of suicide in the post 40 age group where infirmities and health impairment are most common. One potential source of this discrepancy is the overwhelming sense of tragedy assigned to the death of a young person in our culture. This fact may cause news about teenage suicides to travel faster and farther throughout acquaintance networks. However, this reasoning would also lead us to expect a higher percentage of young homicide victims to be reported, a trend present only slightly in the GSS data.

Another potential source of the discrepancy between the Vital Statistics' and the GSS' age distribution of suicide victims are the discrepancies between the social labelling of causes of death as processed by the medical establishment and those "understood" by the population at large (Scheff, 1975; Gove, 1976; Douglas, 1967). A suicide is often understood by bystanders as either the product of a deranged mind or a particular moment of irrationality. Therefore, in order for a suicide to be accepted by a bystander it must first overcome prior conceptions of the victim as well as meeting other evidentiary considerations -- belief is dependent on whether the victim "was the kind of person who would ... ." The family and friends of an elderly, terminally ill person who decides to end his or her life may consider the death a voluntary choice, and subsequently not refer to it as a "suicide."

Table 3 presents cumulative AIDS data from CDC official statistics (1988, 1989, 1990) and from the GSS 1988, 1989, and 1990 surveys.

Insert Table 3 about here.

Increasing percentages of GSS samples over time reported knowing one or more persons with AIDS. All three GSS surveys closely reproduce the CDC-reported exceptionally strong gender imbalance in the disease. With respect to age, all three surveys identify substantially larger proportions of persons with AIDS between 21 and 40 than the CDC reports.

In comparisons with respect to minority status and region, there are sharp departures.[1] The GSS data from all three years suggest that the white proportion of cases is substantially higher than the CDC reports and that a substantially greater share of the cases is in the midwestern region.[2] In

[1]Although it would be desirable to construct confidence intervals to assess the statistical significance of these observed differences, our lack of knowledge about the sampling distribution of the set of PWAs known to GSS respondents makes it difficult to do so. We thus chose not to try to make such estimates because they are likely to be misleading or wrong.

[2]There are a number of surveys that have asked whether respondents knew someone with AIDS. One such study was a telephone survey using random digit dialing done in Chicago from April to July 1987 (Ostrow, Michaels, and Albrecht, 1988). We compared the results from this telephone survey to the official reports on AIDS cases collected by the Chicago Department of Health. The result is strikingly similar to the national comparison: the survey data indicate a higher percentage of white cases than the official statistics (approximately 68 versus 58%) and a lower percentage of black cases (18 versus 33%).

contrast, the East has almost half as many cases proportionally, according to the three GSS surveys, than the CDC reports. The discrepancy between the CDC and GSS figures for the Midwest has been fairly consistent from 1988 through 1990, while the gap has moderated in the case of the South.[1]

One potential explanation for this regional discrepancy is differing levels of sensitivity within the CDC's monitoring system. As we have argued above, both the likelihood that a case will be reported and the length of time necessary to process that case are functions of the alertness of medical professionals to the disease and the efficiency of the bureaucratic reporting channels. This sensitivity of the monitoring system should increase with increasing prevalence of the disease and with sustained efforts on the part of health officials. Therefore, the elevated CDC counts in the East may be due to the mobilization of the major public health systems in New York and New Jersey.

Another potential explanation for the regional discrepancy is that the cases reported by GSS respondents are a more current sample of AIDS cases than those reported by the CDC during the same year. This reasoning may be illustrated by subtracting the 1988 CDC case counts in the regional breakdowns from the 1989 figures and computing the percentages of cases being contributed from each region to the "new" case total. The South contributed 32% of the new cases reported during the year, which is substantially higher than the accumulated case percentages of 27% and 26% reported for 1988 and 1989, respectively. Similarly, the East contributed 35% of the new cases, continuing the downward trend of 39% to 37% reported for 1988 and 1989 accumulated totals, respectively. The Midwest's contribution of 10% to the new case total

---

[1]Since an individual's circle of acquaintances tends to be geographically concentrated in the vicinity of his or her place of residence and more information about one's acquaintances is likely to be available for the physically proximate, one can make fairly solid inferences about the regional distribution of AIDS cases. We found for both the GSS 1988 and 1989 surveys that respondents in the West were more likely to know persons with AIDS then elsewhere, respondents in the Midwest and South were less likely to know persons with AIDS than elsewhere, and respondents in the East were close to the proportionate expectation. It was quite clear that Easterners were not twice as likely to know someone with AIDS as the CDC estimates would imply.

continues the upward move of 8% to 9% in the CDC figures for 1988 and 1989.[1] Because of the immediacy of accessing information in personal networks about newly emerging AIDS cases, network techniques might provide a more responsive means to track shifts in spatial and social distributions of case loads. Although health officials may attempt to speed up the case reporting process, the stringency of bureaucratic routines to insure accuracy and reliability will always cost time in the processing of cases.

It is with respect to racial status that the CDC and GSS estimates are at greatest odds. Both black and Hispanic percentages are substantially lower in the GSS survey estimates than the CDC reports. We noted above that the GSS figures on homicides for 1988 and 1989 reflected a much larger proportion white (and a correspondingly lower proportion black) then the official statistics, but this discrepancy disappears in the 1990 data. Even if we were to adjust the GSS figures on the racial mix of AIDS cases by the discrepancy in the proportion of black homicide victims, there would still be a dramatically lower proportion black among AIDS cases reported by the GSS than the CDC reports indicate.[2] This difference becomes even more apparent by calculating estimates of the racial breakdown for the new cases being reported to the CDC between March 1988 and March

---

[1]While these increases in the proportions of new cases being reported from the South and Midwest may reflect a shift in occurrence of new AIDS cases, they may also result from the monitoring systems in these regions becoming more sensitive while those in the East are becoming fatigued.

[2]If we assume that the official homicide data are much less subject to distortion than the official AIDS data, we can then construct an adjustment for any potential under or overestimation of black vs. non-black homicides by the GSS. This adjustment can then be applied to the GSS AIDS data to refine its estimate of the percentage of black vs. non-black cases.

We chose as our fundamental comparison the ratio of blacks to non-blacks (as opposed to, for example, whites vs. blacks) because the various sources of data use different approaches to counting Hispanics. The official crime statistics sources treat Hispanic as a separate variable independent of racial classification, whereas the CDC and the GSS treat Hispanic as a mutually exclusive category of a race/ethnicity variable. Most Hispanics would classify themselves and be classified by others as white, though a small percentage (about 3%) might be considered black. Thus, black is the most stable category in these two systems (Denton and Massey, 1989).

Since we are using several different data sources, combining estimates seems to be a reasonable approach. Thus, we calculate a combined estimate of the ratio of the proportion of blacks to non-blacks from the official statistics using both the Vital Statistics and the UCR of .88 (= 18,801/21,256 = .47/.53), while a combined estimate from the three years of the GSS is .67 (= 277/413 = .40/.60). Therefore, the resulting adjustment factor for the ratio of blacks to non-blacks for the GSS homicide figures is .88/.67 = 1.32. For the 1990 GSS AIDS estimates, the raw ratio of blacks to non-blacks is .12/.88 = .14. If we apply the adjustment derived from the homicide data this becomes 1.32 x .14 = .18. That is, the adjusted ratio of percent of black to percent of non-black cases is about .18. The CDC figure of 35,472 blacks implies a ratio of .38, which is over double the adjusted GSS figure. Even if we only use GSS homicide data from 1988 and 1989, which produces a larger adjustment of 1.52, we get an adjusted black to non-black ratio for AIDS from the GSS of .21, still close to one-half the CDC figure.

1989 and between March 1989 and March 1990. The CDC figures suggest a declining share of white cases -- 52 percent of the new cases in 1989-90 are reported as white whereas 57 percent of the previously reported cases were white. There are corresponding increases in the black and Hispanic percentages. The GSS estimates reflect an apparent reversal of the CDC patterns. Given the small sample sizes, we cannot claim that the GSS estimates indicate a trend toward a greater white share of the AIDS cases, and we do not suggest such a projection. But there is clearly an important anomaly in these two sets of data that deserves attention.[1]

This anomaly may be largely due to differential coverage by the two methods. For example, an analysis of the association between exposure category and race reveals that white AIDS cases are predominantly men who have sex with men, while among minority populations there are clearly two different exposure groups: men who have sex with men and men who use intravenous drugs. Many of the white men who have sex with men are gay identified. However, black and Hispanic men who have sex with men often conceal their sexual interests and life style as well as the fact that they have AIDS from their peers.[2] In addition, the intravenous drug users are often isolated from others in the minority community. The result is that both minority groups may be lost to the scanning practices of their peers and therefore undercounted when using network surveillance techniques.

Similarly, the recent GAO report on undercounting in the official registry of AIDS cases suggests that the undercount results from a variety of sources and may be as high as 35 to 41 percent. Some of the sources identified by the GAO report are: HIV infections that do not officially qualify as AIDS; the failure to diagnose cases; presumptive diagnoses that are never accepted or accepted late by the CDC; and the failure to report properly diagnosed cases. It is

---

[1]Black respondents in the GSS samples are, in fact, more likely to know someone with AIDS than white respondents are, although white college graduates are much more likely to know someone with AIDS than blacks are.

[2]It is interesting to note that people claiming to know persons with AIDS typically acknowledge their social ties as being casually acquainted rather than socially close, while persons claiming to know a homicide or suicide victim more often acknowledge the victim as being more closely tied socially to the reporter. This suggests that people may find it easier to acknowledge a socially stigmatized person if he is socially distant than if he is a member of one's own intimate circle of friends and kin. This effect may also contribute to an underreporting of minority AIDS cases by the GSS, since AIDS is often more stigmatized in minority communities.

unclear how HIV infections that do not officially qualify as AIDS would differentially affect reporting to the CDC or in respondent networks. However, given that the original disease was most strongly characterized among white gay men, manifestations of HIV disease among intravenous drug users and women might less often be counted as qualifying as AIDS. The failure to officially diagnose cases might arise from three causes: self-conscious medical decisions not to label persons as having AIDS, the failure to identify AIDS among a variety of other conditions, or the failure to identify AIDS among persons who are believed to be socially unlikely to have the disease. All of these would contribute to an undercount of white middle class persons, cases in low incidence areas, and groups with low incidences of cases while only the second would produce an undercount of persons in metropolitan areas whose health status involved many diseases (e.g., IVDUs). Presumptive diagnoses, based on knowledge of an experienced physician that he or she is dealing with AIDS may reduce the number of minority cases accepted by the CDC from overworked metropolitan hospitals. Finally, the failure to report properly diagnosed cases may be differential across regions and may be more common in overburdened reporting systems at the epicenters of the epidemic. The differential impact of this problem on differential case reporting by risk group or socio-demographic group is unclear.

If the official statistics do not record substantial numbers of cases, it is likely that some proportion of the differences between the network data and the CDC reports rest upon each method capturing cases that the other does not. The network method may less efficiently capture cases from minority communities and highly concentrated urban epicenters (both of gay men and IVDUs); on the other hand, it may be more efficient in areas where cases are underreported to the official system, where cases are less common, and among the socially more advantaged.

Trends Over Time[1]

In order to examine trends in the proportion of persons who know a PWA over time, we used the GSS data in conjunction with data from the National Health Interview Survey (NHIS) from August 1987 to March 1990. The NHIS is a continuous, monthly cross-sectional household interview survey drawn from independent probability samples of the civilian non-institutionalized population (National Center for Health Statistics, 1988-90). The data used here, from the NHIS supplement on AIDS knowledge and attitudes, are from interviews of one randomly chosen adult 18 years or older in each family. The response rates vary from 80 to 89 percent of eligible respondents. The targeted population, namely adults living in households, is the same for both the GSS and the NHIS. The NHIS achieves a slightly higher response rate, but on the basis of the brief descriptions provided, the samples and basic interview technique appear to be comparable. Both surveys ask almost identical questions to determine if the respondent has known someone with AIDS.

Figure 1 shows the percentage of respondents who personally know at least one PWA plotted against the number of cumulative cases reported to the CDC during the month in which each survey was administered.

Insert Figure 1 about here.

Because the official number of cumulative AIDS cases has risen linearly[2] across each monthly interval, each point on the graph represents, in chronological order, an individual NHIS or GSS survey. Approximate 95% confidence intervals are shown for each point. These error bars permit a visual assessment of the statistical significance of comparisons between the individual point estimates. The proportions are generally increasing with increased cumulative cases (or time),

[1]The analyses reported in this section draw heavily on similar analyses completed by Patricia Styer using GSS and NIHS data collected through March, 1989.

[2]A scatterplot of time versus cumulative cases reveals an almost perfect linear relationship between the two. The only departure from linearity is a slight autocorrelation among the residuals, a pattern that is common in time series data such as these.

indicating a strong link between the cases reported to the CDC and those picked up through the GSS network question.

If there were not a strong link between the cumulative cases reported by the CDC and the percentage of respondents who report knowing a PWA, comparisons between our network estimates of AIDS rates and those of the CDC would be meaningless. Yet this fact alone is uninteresting. What is interesting from a social epidemiological perspective is the character of the numerical relationship between the proportion knowing a PWA and the number of cumulative cases. If AIDS cases occurred randomly throughout the population, and if the relevant information about a given victim was communicated to and evaluated by that victim's acquaintances in the same way physicians diagnose and report the disease among their patients, then we might expect a perfect correspondence between increases in cumulative cases and increases in proportions of persons knowing PWAs. The extent to which the empirical relationship between these two sets of numbers changes over time indicates the nonrandom occurrence of the disease and possible changes in the way people communicate and evaluate information about possible AIDS victims.

To assess whether the relationship between the proportion who know a PWA and cumulative cases has changed over time, we must make some judgement about the functional form of the relationship depicted in Figure 1. The proportions are steadily increasing, but at a decreasing rate. In an attempt to model this trend, we began by fitting the logistic regression of cumulative cases on the proportion of people who know a PWA, using each individual NHIS or GSS survey as a single observation:

$$\text{logit}(\theta) = \mu + \beta(\text{cumulative cases}) + e \qquad (1)$$

where θ represents an individual's probability of knowing someone with AIDS, and the error terms are distributed binomially. This regression models the association between cumulative cases and the proportion who know a PWA as non-linear and non-additive, weighting each observation by the sample size of that particular survey. Partitioning the likelihood ratio statistic (G^2) and comparing the part accounted for by cumulative cases to a χ^2 distribution on one degree of freedom allows us to test the fit of Model 1 against the null hypothesis in which the logit(θ) depends only on the mean parameter μ. This test clearly demonstrates that the effect of cumulative cases is significant ($G^2 = 597$ on 1 degree of freedom). However, the overall fit of the model is not adequate ($G^2 = 97$ on $21 - 2 = 19$ degrees of freedom). An examination of the residuals plotted against the fitted values (Figure 2) reveals that the model overestimates the proportions at the extremes, while underestimating the proportions near the middle.

Insert Figure 2 about here.

Because this curvature of the residuals relative to the fitted values implies a semi-logarithmic form, we transformed cumulative cases using logarithms and estimated a new regression. This new model, plotted in Figure 3, allows for the slope of the fitted line to decrease as the cumulative number of AIDS cases increases.

Insert Figure 3 about here.

As in Model 1 the partial G^2 evaluating the effect of cumulative cases (log transformed) is highly significant ($G^2 = 632$ on 1 degree of freedom). Unfortunately, the new model still does not provide an adequate fit to the data ($G^2 = 62$ on 19 degrees of freedom). Even the removal of the three GSS observations does not result in an adequate fit, although it does reduce G^2 somewhat (45 on 16 degrees of freedom). This persistent lack of fit of the model does not necessarily imply that the new functional form we have specified is incorrect -- only that there are additional components of variation within the data that cannot be explained by the steady increase in cumulative cases.

Since neither Model 1 nor the transformed model is a special case of the other, we are unable to partition G^2 to show that the model with the log transformed predictor fits significantly better than the simple logistic model. However, transforming the predictor does reduce the overall G^2 and provides a qualitative improvement in fit, as demonstrated by a comparison of the two residual plots in Figure 2. The last four NHIS surveys, due to their large sample size, are highly influential in determining the parameter estimates of the model. These points, along with the '90 GSS survey (together representing 9 months), clearly indicate a levelling off of the growth in the percentage of the population who know a PWA. This may be reflective of the embeddedness of the acquaintance networks of AIDS victims. If a majority of victims and their acquaintances belong to a distinct set of social groups (e.g., IV drug users, gay men, and persons with many sexual partners), these groups may have become sufficiently saturated with PWAs that additional victims do not produce a commensurate increase in the overall percentage of the population who know someone with AIDS.[1]

Note that the GSS '89 and '90 proportions both appear to be lower than the neighboring NHIS proportions. It is possible that the difference between the GSS '90 point and the preceding NHIS point occurred by chance -- a 95% confidence interval around the GSS '90 point includes not only the preceding NHIS point but also its entire confidence interval. However, it is also possible that the GSS question is measuring something slightly different than the NHIS question. The GSS question is asked after a variety of questions unrelated to AIDS. In contrast, the NHIS question occurs at the end of a large set of questions about AIDS knowledge and beliefs. It is possible that this prior questioning primes NHIS respondents to identify PWAs among their set of acquaintances. Although our evidence is not sufficient to prove that this effect exists, the issue is important in refining and improving a network monitoring strategy.

Because AIDS is mainly transmitted through direct social contact (such as sexual intercourse or needle sharing), the spread of AIDS has been channelled by the social structure of the society. This fact is revealed by persistent differences in the prevalence of and increase in AIDS cases across socially relevant population subgroups. However as we have argued above, the CDC figures may provide an imprecise view of the relative prevalence of AIDS cases across population subgroups defined by race, age, and geographical region. In contrast, the AIDS victims known to survey respondents may be more reflective of the true composition of AIDS

[1]This interpretation is also supported by the distribution of the actual number of PWAs known to respondents. If knowing a PWA were indeed a random occurrence, we would expect far more people to know one PWA and far fewer to know three, four, and five PWAs than what actually is observed.

cases in the population. Moreover, additional variables that are important in structuring social interaction such as level of education are not reported by the CDC.

To examine differences in the increase of AIDS cases across subgroups requires more individual estimates than the three GSS surveys. Unfortunately, the NHIS surveys do not collect information about the PWAs themselves. However for those subgroups defined by race, level of education, and to a lesser extent age, we expect the majority of respondents' acquaintances to fall within their own subgroup. This expectation is confirmed by numerous studies indicating a strong association between the social characteristics of individuals and of their associates. In addition, almost all of the GSS respondents who reported knowing PWAs also reported that those PWAs had the same characteristics as themselves.

Using this expectation about individuals' acquaintances, we propose the proportion of respondents in a specific subgroup who know a PWA to be a rough indicator of the number of actual AIDS cases within that subgroup. Moreover, the change over time in the proportion of that subgroup who know a PWA is an indicator of the historical spread of AIDS within that subgroup. To measure these relationships, we fit the same logistic model as in Figure 3 (Model 1 with the log transform of cumulative cases) to subgroups of respondents based on age, race, and level of education.[1] The parameter estimates and goodness of fit statistics are presented in Table 4.

Insert Table 4 about here.

Differences in the parameter estimates across these social groups may be indicative of differences in the prevalences and rates of increase of AIDS cases within the groups.

With respect to age, the logistic regressions show a significant increase over time in the proportion of respondents who know a PWA for each age group. The slope coefficients, although small, are all significant as judged by either a chi-squared statistic or t-statistic. Although there are slight differences in the slopes among the three groups, these differences are not statistically

---

[1]We were unable to examine the increase in AIDS within regional subgroups because tabulations of the NHIS data by region have not been published nor are the recent data publicly available for analysis.

significant. The highest intercept was for the 30-49 age group, while the lowest intercept was for the 50+ age group.[1] These models suggest that, although the number of AIDS cases in each age group have increased over the past three years, the majority of initial cases were within the 30-49 age group and have remained there.

The logistic regressions for each educational group also showed a significant increase over time in the proportion of respondents who know a PWA. Each group begins at roughly the same level (similar intercept estimates), however the slope estimates increase as level of education increases. While the slope for high school graduates is not statistically different from that for respondents with less than 12 years of education, the slope for those with at least some college is significantly larger than both of the former. These results suggest not only that AIDS is concentrated among highly educated persons, but that its concentration has increased substantially during the period under analysis. This increase in the positive relationship[2] between educational attainment and knowing someone with AIDS underscores the fact that the modes of transmission of the disease are socially organized.

For both blacks and whites, there is a significant increase in the proportion of respondents knowing a PWA, though the evidence for blacks is considerably weaker. Although the intercept for blacks is significantly higher than that for whites, the slope for whites is higher (though not significantly), indicating a possible shift in the distribution of the disease between the two groups.

We believe that these results are important clues to the changing distributions of AIDS among population subgroups, subject to our assumptions about the homogeneous composition of

[1] Although there is a distinct age difference in the likelihood of knowing a person with AIDS, there is no sex difference. Why is the network procedure, resting, as it does, on the logic of social networks, sensitive to age but not to sex differences among respondents? A plausible explanation is to consider the social composition of the respondents' personal acquaintanceship networks. We know from other studies (Laumann, 1973; Fischer, 1982; Marsden, 1982; Burt, 1986) that age plays a highly selective role in determining one's acquaintances, but that gender plays a much less selective role -- that is, there is a rough parity in the likelihoods of individuals of either sex knowing men and women.

[2] In fact, higher status persons are much more likely to know persons with AIDS than lower status respondents (e.g., 1989 GSS respondents with postgraduate degrees are 3.2 times as likely to know a person with AIDS than are high school graduates). Interestingly, three indicators of the respondents' socioeconomic standing (occupational prestige, subjective class identification, and education attainment) all suggest that higher status respondents are less likely to know a homicide victim than are lower status respondents. This strong reversal of the trend for knowing persons with AIDS is consistent with what is known about homicide victimization generally (Wolfgang, 1958; Braucht et al., 1980; Nettler, 1982).

respondents' acquaintance nets. However, regardless of the accuracy of these models in reflecting the actual distribution of AIDS cases, they do describe accurately the differential likelihood of knowing a PWA for population subgroups. If knowing a PWA affects one's attitudes both about AIDS and about groups of people who have been associated with the disease, then our results are indicative of where political and social support for efforts to combat the disease are most likely to come from.

Conclusion

In this paper we have proposed a method for making independent estimates of the relative prevalence of specific health events across population subgroups. We illustrated the technique using AIDS, homicide, and suicide, because these events have similar incidences in the general population, and because each has a different social epidemiology and a unique set of social meanings attached to it. Our network technique reproduces many of the features of the annual incidence of homicide and suicide and the relative prevalence of AIDS among males and females, as reported by the official agencies. While this correspondence between the official figures and the network estimates does not prove that our network technique is always accurate, it does suggest that the technique can provide good approximations with a sample size of only 1500.

In addition to this correspondence with the official figures, the network estimates are quite consistent across time. The GSS '88, '89, and '90 samples all yield similar relative percentages across subgroups of both PWAs and homicide victims. The NHIS estimates of the proportion of people who know a PWA have also increased consistently with the total number of AIDS cases reported by the CDC. This consistency of our proposed network method suggests that results based on this method are stable, and that any biases present in the approach are also stable.

The advantage of figures reported by official agencies is that they represent a steady and unforgotten accumulation of cases because bureaucratic memories tend to be formalized and systematic and thus reliable. Therefore, official statistics may provide the most stable, if not the most accurate descriptions of events that have occurred in the past. In addition, official statistics are at this time the only estimates of the total number of persons affected by specific health events. These total numbers are critical in assessing the present and future impact of an event, and in formulating effective policies to address it.

In contrast, the strength of a network monitoring system is in the timeliness of the information that it gathers. Information about a PWA is probably transmitted more quickly through an informal acquaintance network than through the bureaucratic channels in a formal case reporting hierarchy. Moreover, the network technique is likely to respond more quickly to changes in the set of criteria sufficient to diagnose an event than are the official reporting systems. The result is a monitoring device that is more responsive to the distribution of recently emerging cases, giving us a preview of coming developments. Finally, because official reporting systems attempt to record every occurring case, they are limited (sometimes by bureaucratic practice) in the amount of information that they can collect about the affected persons. A network monitoring system could enhance that information by adding additional items to its questionnaire. In sum, a network monitoring system is valuable in its ability to provide current information about an event, to indicate specific areas in which the official reporting system might be weak, and to provide additional information about affected persons in a cost-effective way.

The analyses presented here are not intended to be decisive but rather illustrative of the reasoning behind, the potential uses of, and the problems involved in a network monitoring system. Clearly a more comprehensive research program is necessary to develop the technique further. We see the following issues as important topics for future research. First, AIDS, homicide, and suicide are only a few of a large number of relatively rare, widely distributed, yet stigmatized phenomena that societies need to measure. Other examples include abortion, rape, and family violence. Using the network monitoring technique to measure several different events might help identify the strengths and weaknesses of the official systems responsible for reporting those events and provide a clearer picture of the biases that affect network monitoring.

Not all rare events can be monitored using the network technique. For example, the event must be of the nature that information about it is communicated to acquaintances. Therefore, the incidence of certain sexual behaviors cannot be measured in this way. In addition, the respondent must be willing to reveal information about his or her acquaintances relevant to the event. For example, respondents may have reservations about revealing the fact that an acquaintance has committed a crime for fear that they will be forced to identify that acquaintance. The result is that criminal behaviors can probably only be measured through victims. Thus, "victimless" crimes such as tax evasion, theft, and drunk driving would probably not be effectively measured using the network technique. Finally, the event must be socially defined and recognized by a majority of the population. For this reason, it would not be useful to ask people if they knew of anyone who came from a "dysfunctional family."

A second topic for further research concerns the potential biases in the network technique. While some of these biases are caused by the social significance of the event being measured, others result from social processes independent of the event. For example, systematic differences in the size of acquaintance nets across subgroups might yield a biased set of acquaintances affected by an event. One method for estimating the effect of these biases would be to ask people to scan their acquaintance nets for the presence of an event that occurs randomly in the population and to which no social stigma is attached, such as the occurrence of twins. The results from this question, when compared with what would be expected, could be used to correct for biases in responses to questions about other events.

The third issue to be addressed is that of using the results from the network monitoring questions to make independent estimates of the total number of affected persons in the population. One approach would be to take an estimate of the average number of acquaintances known to an individual and multiply this number by the GSS sample size to get an estimate of the total number of acquaintances scanned by the respondents. This total could then be divided into the total number of PWAs reported, yielding an estimate of the proportion of the population with AIDS. Unfortunately there are two problems with this approach. Because the signal strength of information about different events is likely to vary, a separate estimate of the total number of acquaintances scanned would have to be made for each event. More importantly, the occurrence of events such as AIDS is not distributed randomly throughout the population. Instead, the likelihood of knowing a PWA is different for different people. These differences would have to be unravelled before any estimates of the total number of AIDS cases could be made.

Future research into these and other topics aimed at refining the network monitoring technique should yield a new and powerful tool for exploring the social epidemiology of rare health events. Understanding the social epidemiology of an event such as AIDS is critical in targeting efforts to combat the disease and in developing projections about its future impact. Such understanding will also help to identify subgroups of affected persons who are being missed by the official reporting systems. As an added bonus, attempts to model respondents' answers to network questions like those discussed here will provide social network researchers with information about acquaintance networks and the flow of information through those networks.

References

Altham, P. M. E. 1976. "Discrete Variable Analysis for Individuals Grouped into Families." *Biometrika* 63:263-9.

Berkowitz, Stephen D. 1982. *An Introduction to Structural Analysis: The Network Approach to Social Research.* Toronto: Butterworths.

Bernard, H. Russell, Eugene C. Johnsen, and Peter D. Killworth. 1988. "Estimating the Size of Hard-To-Count Populations." Grant proposal to the National Science Foundation, Washington, D.C.

Bernard, H. Russell, Eugene C. Johnsen, Peter D. Killworth, and Scott Robinson. 1987. "Estimating the Size of an Average Personal Network and of an Event Subpopulation: Some Empirical Results." *Proceedings of the Section on Survey Research Methods, American Statistical Association.* Appendix C: pp. 1-9.

Boissevain, J. 1974. *Friends of Friends: Networks, Manipulations, and Coalitions*. Oxford: Blackwell.

Braucht, G., et al. 1980. "Victims of Violent Death: A Critical Review." *Psychology Bulletin* 87:309-33.

Brier, Stephen S. 1980. "Analysis of Contingency Tables Under Cluster Sampling." *Biometrika* 67:591-6.

Bureau of the Census. 1983. *1980 Census Population: General Population Characteristics. Part I, United States Summary*. Washington, D.C.: Government Printing Office.

Burt, Ronald S. 1986. "A Note on Sociometric Order in the General Social Survey Network Data." *Social Networks* 8:149-74.

Burt, Ronald S. 1980. "Models of Network Structure." Pp. 79-141 in *Annual Review of Sociology*. Palo Alto, CA: Annual Reviews Inc.

Burt, Ronald S. and Michael Minor (eds.). 1983. *Applied Network Analysis*. CA: Sage.

Centers for Disease Control. March 1988. *AIDS Weekly Surveillance Reports*. Atlanta.

Centers for Disease Control. March 1989. *HIV/AIDS Surveillance Report*. Atlanta.

Centers for Disease Control. March 1990. *HIV/AIDS Surveillance Report*. Atlanta.

Cohen, J. E. 1976. "The Distribution of the Chi-Squared Statistic Under Clustered Sampling from Contingency Tables." *Journal of the American Statistical Association* 71:665-70.

Conway, George A., Brenda Colley-Niemeyer, Claire Pursley, Carlota Cruz, Sherin Burt, Pam Rion, and Clark W. Heath, Jr. 1989. "Underreporting of AIDS Cases in South Carolina, 1986 and 1987." *Journal of the American Medical Association* 262:2859-63.

Curran, James W., Harold W. Jaffe, Ann M. Hardy, W. Meade Morgan, Richard M. Selik, and Timothy J. Dondero. 1988. "Epidemiology of HIV Infection and AIDS in the United States." *Science* 239:610-6.

Current Population Reports, Series P-25, No. 1024. 1988. *State Population and Household Estimates with Age, Sex, and Components of Change: 1981-87*. Washington, D.C.: Government Printing Office.

Davis, James A. and Smith, Tom W. 1989. *General Social Surveys, 1972-1989: Cumulative Codebook*. Chicago: National Opinion Research Center.

Denton, Nancy and Douglas Massey. 1989. "Racial Identity Among Caribbean Hispanics: The Effect of Double Minority Status on Residential Segregation." *American Sociological Review* 54:790-808.

Dondero, Timothy J., M. Pappaioanou and James W. Curran. 1988. "Monitoring the Levels and Trends of HIV Infection: The Public Health Services HIV Surveillance Program." *Public Health Report* 103:213-20.

Douglas, Jack D. 1967. *The Social Meanings of Suicide*. Princeton: Princeton University Press.

Feinstein, Alvan R. 1985. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia: W. B. Saunders Company.

Fischer, C. 1982. *To Dwell Among Friends*. Chicago: University of Chicago Press.

Freeman, Linton C. and C. R. Thompson. 1989. "Estimating Acquaintance Volume." in *The Small World*, edited by M. Kochen. Norwood, NJ: Ablex.

GAO. 1989. *AIDS Forecasting: Undercount of Cases and Lack of Key Data Weaken Existing Estimates*. Washington, D.C.: GAO/FMED-89-13.

Gove, Walter (ed.). 1976. *The Labelling of Deviance*. Wiley.

Gove, Walter and M. Geerken. 1985. "Are Uniform Crime Reports A Valid Indicator of the Index Crimes? An Affirmative Answer with Minor Qualifications." *Criminology* 23:451-501.

Holland, P. and S. Leinhardt (eds.). 1979. *Perspectives on Social Network Research*. New York: Academic Press.

Institute of Medicine. 1986. *Confronting AIDS: Directions for Public Health, Health Care, and Research*. Washington, D.C.: National Academy Press.

Kalton, Graham and Dallas W. Anderson. 1986. "Sampling Rare Populations." *Journal of the Royal Statistical Society* 149:65-82.

Laumann, Edward O. 1973. *Bonds of Pluralism*. New York: Wiley.

Laumann, Edward O., John H. Gagnon, Stuart Michaels, Robert T. Michael, and James S. Coleman. 1989. "Monitoring the AIDS Epidemic: A Network Approach." *Science* 244:1186.

Laumann, Edward O. and David Knoke. 1987. *The Organizational State: Social Choice in National Policy Domains*. Madison: University of Wisconsin Press.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52:122-131.

Mitchell, J. C. (ed.). 1969. *Social Networks in Urban Situations*. Manchester, England: Manchester University Press.

Modesitt, Steven K., Sharon Hulman, and David Fleming. 1990. Evaluation of Active Versus Passive AIDS Surveillance in Oregon." *American Journal of Public Health* 80:463-4.

National Center for Health Statistics. 1988. *Vital Statistics of the United States, 1986, Vol. II, Mortality, Part B*. Washington, D.C.: Government Printing Office.

National Center for Health Statistics. 1989. *Vital Statistics of the United States, 1987, Vol. II, Mortality, Part B*. Washington, D.C.: Government Printing Office.

National Center for Health Statistics. 1988-90. *Advance Data*. Nos. 146, 148, 150, 151, 153, 160, 161, 163, 164, 167, 175, 176, 179, 183, 186, 193.

Nettler, G. 1982. *Killing One Another*. Cincinnati: Anderson.

Nettler, G. 1978. *Explaining Crime*. New York: McGraw-Hill.

Ostrow, D. G., Stuart Michaels, and G. A. Albrecht. 1988. "Information and Misinformation: The State of Knowledge, Attitudes, and Beliefs about AIDS in the Chicago Metropolitan Area General Population." Preliminary Report to the Chicago Department of Health's Comprehensive AIDS Prevention Education Program. Chicago.

Pool, I. de Sola and M. Kochen. 1978. "Contacts and Influence." *Social Networks* 1:5-51.

Scheff, T. J. (ed.). 1975. *Labelling Madness*. New Jersey: Prentice Hall.

Shilts, R. 1987. *And The Band Played On: Politics, People, and the AIDS Epidemic*. New York: St. Martin's Press.

Sirken, Monroe G. 1970. "Household Surveys with Multiplicity." *Journal of the American Statistical Association* 65:257-266.

Sirken, Monroe G. 1972. "Stratified Sample Surveys with Multiplicity." *Journal of the American Statistical Association* 67:224-27.

Sudman, Seymour and Graham Kalton. 1986. "New Developments in the Sampling of Special Populations." Pp. 401-29 in *Annual Review of Sociology*. Palo Alto, CA: Annual Reviews Inc.

Sutherland, E. and D. Cressey. 1978. *Criminology*. Chicago: Lippincott.

Traugott, Michael W., Robert M. Groves, and Theodore Downes-Le Guin. 1990. "Issues in the Measurement of Contact With Persons With AIDS." Unpublished manuscript. Institute for Social Research, University of Michigan.

Turner, C. F., H. G. Miller, and L. E. Mosses (eds.). 1989. *AIDS: Sexual Behavior and Intravenous Drug Use*. Washington, D.C.: National Academy Press.

U.S. Department of Justice, Federal Bureau of Investigation. 1989. *Uniform Crime Reports, 1988*. Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Justice, Federal Bureau of Investigation. 1990. *Uniform Crime Reports, 1989*. Washington, D.C.: U.S. Government Printing Office.

Wellman, Barry and Stephen D. Berkowitz (eds.). 1988. *Social Structures: A Network Approach*. Cambridge, England: Cambridge University Press.

Wilson, William J. 1987. *The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy*. Chicago: University of Chicago Press.

Wolfgang, M. 1958. *Patterns of Criminal Homicide*. Philadelphia: University of Pennsylvania Press.

Table 1: Annual Homicide Data from Official Statistics and GSS 1988 - 90

	UCR 1988		UCR 1989		Vital Stats 1987(a)		GSS 1988(b)		GSS 1989(c)		GSS 1990(d)	
	N	% Dist'n	N	% Dist'n	N	% Dist'n	N	% Dist'n	N	% Dist'n	N	% Dist'n
Sex												
Male	13,632	75	14,464	76	15,855	75	121	79	179	74	146	74
Female	4,611	25	4,483	24	5,248	25	31	20	59	25	49	25
Unknown	26	0	7	0	0	0	2	1	2	1	2	1
Total	18,269	100	18,954	100	21,103	100	154	100	238	100	197	100
Race												
White	9,003	49	9,103	48	11,128	53	142	56	150(e)	63	92(e)	47
Black	8,786	48	9,314	49	9,487	45	95	37	86	36	96	49
Other	301	2	344	2	488	2	18	7	0	0	6	3
Unknown	179	1	193	1	0	0	0	0	2	1	3	2
Total	18,269	100	18,954	100	21,103	100	255	100	238	100	197	100
Ethnic Origin(f)												
Hispanic	2,841	15					23(e)	9	18	8	10	5
Non-Hisp	12,868	67					232	91	218	92	184	93
Unknown	3,548	18					0	0	2	1	3	2
Total	19,257	100					255	100	238	100	197	100
Age(g)												
10 or less	775	4	753	4	748	4	2	1	4	2	2	1
11-20 yrs	1,949	11	2,248	12	2,104	10	26	17	43	18	35	18
21-40 yrs	10,649	58	11,022	58	12,307	58	90	58	147	62	121	61
41 or older	4,638	25	4,674	25	5,843	28	35	23	42	18	37	19
Unknown	308	2	257	1	101	0	1	1	2	1	2	1
Total	18,269	100	18,954	100	21,103	100	154	100	238	100	197	100
Region												
East	3,811	18	3,972	18	3,456	16	41	16	33	14	32	16
Midwest	3,848	19	3,990	18	4,110	19	70	27	54	23	44	22
South	8,723	42	9,221	42	9,172	43	106	42	102	43	75	38
West	4,293	21	4,317	20	4,451	21	38	15	40	17	43	22
Other(h)	0	0	467	2	0	0	0	0	0	0	1	1
Unknown	0	0	0	0	0	0	0	0	9	4	2	1
Total	20,675(i)	100	21,967	100	21,189	100	255	100	238	100	197	100

Official statistics are Uniform Crime Reports (UCR) 1988 & 1989 and Vital Statistics 1987. GSS 1988, 1989, and 1990 are the 1988, 1989, and 1990 General Social Surveys (GSS) respectively. Categories of the percent distributions do not always add to 100 due to rounding.

(a) The counts in this column are of all "homicides & legal interventions" (i.e., deaths due to homicide and injury purposely inflicted by other persons including law-enforcement agents acting in the line of duty) from the Vital Statistics for the United States 1987. Ethnic Origin was not reported in the Vital Statistics for 1987.

(b) The data for the GSS 1988 are characterizations of homicide victims (in last year) known to respondents. Respondents gave a specific description of their relation to and the demographic characteristics only of the victim they knew best. Victims whose relation to the respondent was "patient" are not included in these tabulations. Thus, the tabulation for sex and age are based only on the single closest victim. However, race, ethnic origin, and region are based on the total number of victims mentioned. Race and ethnic origin of the closest victim of a homicide come from a single variable with 4 categories: black, white, Hispanic, or other. Race and ethnic origin for victims other than the closest one were assigned the characteristic of the respondent based on a crosstabulation of the total number of victims of homicide known, the race of the closest victim, and respondent's race (ethnic origin). Region of homicide was not asked in 1988 and all homicide victims were assigned the region of the respondent.

Table 1 (cont.)

(c) In the 1989 GSS, respondents were asked about the demographics of all victims of homicide they had known in the past 12 months. Victims whose relation to the respondent was "patient" are not included in these tabulations. One respondent reported on multiple victims of homicide which occurred outside the United States and they are not included in these tabulations.

(d) In the 1990 GSS, 3 victims whose relation to the respondent was "patient" are not included in these tabulations.

(e) In 1989 and 1990, respondents were asked about the race/ethnicity of each victim they knew. Hispanic respondents are combined with the white racial category.

(f) Ethnic origin was not reported in the Uniform Crime Reports for 1988 or 1989, therefore the reports from 1986 are used instead. Since the Uniform Crime Reports code race and ethnic origin separately, we reassigned homicide victims identified as Hispanic in the 1988 GSS to one of the three racial categories by looking at both how the victim was identified, respondent's race, and respondent's national origin.

(g) The age breakdown listed is from the GSS. The closest categories available from the official statistics are: less than 10 years old, 10 to 19 years; 20 to 39 years; and 40 years and older.

(h) The Other category is made up of Puerto Rico, the Virgin Islands, Guam, and the Trust Territory.

(i) This total, 20,675, is different than the total in the sex, race, ethnic origin, and age breakdowns. It is considered the correct total number. The more specific information is only available on a smaller number of cases.

Table 2: Annual Suicide Data from Official Statistics and GSS 1990

	Vital Statistics 1986		Vital Statistics 1987		GSS 1990	
	N	% Dist'n	N	% Dist'n	N	% Dist'n
Sex						
Male	24,226	78	24,272	79	124	77
Female	6,678	22	6,524	21	36	22
Unknown	0	0	0	0	1	1
Total	30,904	100	30,796	100	161	100
Race(d)						
White	28,437	92	28,217	92	130	81
Black	1,892	6	1,963	6	20	12
Hispanic(a)	na	na	na	na	8	5
Other	575	2	616	2	2	1
Unknown	0	0	0	0	1	1
Total	30,904	100	30,796	100	161	100
Age						
10 or less	5	0	1	0	0	0
11-20 yrs	2,146	7	2,152	7	31	19
21-40 yrs	12,787	41	12,476	41	78	48
41 or older	15,954	52	16,154	52	51	32
Unknown	12	0	13	0	1	1
Total	30,904	100	30,796	100	161	100
Region						
East	4,648	15	4,711	15	33	20
Midwest	7,218	23	7,244	23	39	24
South	11,329	37	11,216	36	58	36
West	7,755	25	7,669	25	29	18
Other	0	0	0	0	2	1
Total	30,950	100	30,840	100	161	100

Official statistics are Vital Statistics 1986 and 1987. GSS 1990 is the General Social Survey (GSS).
Categories of the percent distributions so not always add to 100 due to rounding.
(a) Ethnic origin was not reported in the Vital Statistics in 1986 and 1987.

Table 3: Cumulative AIDS Data from Official Statistics and GSS 1988 - 90

	CDC Weekly Report March 1988		CDC Weekly Report March 1989		CDC Weekly Report March 1990		GSS 1988(a)		GSS 1989(b)		GSS 1990(c)	
	N	% Dist'n	N	% Dist'n	N	% Dist'n	N	% Dist'n	N	% Dist'n	N	% Dist'n
Sex												
Male	50,647	92	79,918	91	115,575	90	126	95	195	88	230	90
Female	4,520	8	8,178	9	12,744	10	5	4	16	7	17	7
Unknown	0	0	0	0	0	0	1	1	10	5	9	4
Total	55,167	100	88,096	100	128,319	100	132	100	221	100	256	100
Race(d)												
White	32,999	60	50,462	57	71,517	56	167(e)	72	162	73	189	74
Black	14,089	26	23,518	27	35,472	28	43	18	31	14	31	12
Hispanic	7,575	14	13,300	15	20,062	16	13	6	19	9	23	9
Other	504	1	617	1	953	1	9	4	1	0	4	2
Unknown	0	0	99	0	315	0	1	0	8	4	9	4
Total	55,167	100	88,096	100	128,319	100	233	100	221	100	256	100
Age(f)												
10 or less	886	2	1,440	2	2,192	2	0	0	0	0	2	1
11-20 yrs	234	0	352	0	500	0	5	4	7	3	7	3
21-40 yrs	36,990	67	58,903	67	85,301	66	96	73	173	78	194	76
41 or older	17,057	31	27,401	31	40,326	31	30	23	33	15	44	17
Unknown	0	0	0	0	0	0	1	1	8	4	9	4
Total	55,167	100	88,096	100	128,319	100	132	100	221	100	256	100
Region(g)												
East	23,947	39	37,374	36	45,690	36	44	19	45	20	61	24
Midwest	4,868	8	8,670	8	11,057	9	47	20	30	14	41	16
South	15,782	25	28,138	27	36,336	28	58	25	77	35	74	29
West	16,575	27	25,675	25	31,453	25	84	36	57	26	68	27
Other	1,028	2	2,782	3	3,783	3	0	0	0	0	2	1
Unknown	0	0	0	0	0	0	0	0	12	5	10	4
Total	62,200	100	102,621	100	128,319	100	233	100	221	100	256	100

Official AIDS Statistics are from Center for Infectious Diseases, Centers for Disease Control, AIDS Weekly Surveillance Report, March 7, 1988 and HIV/AIDS Surveillance Report, March 1989 and 1990. Categories of the percent distributions do not always add to 100 due to rounding.

(a) Information on sex and age in the GSS 1988 is based on the characteristics of the person with AIDS closest to the respondent only. In 1988, there was no question about region for the person(s) with AIDS; the total number of persons with AIDS known were assigned to the region where the respondent is currently living. In 1988, six respondents reported that the person with AIDS they knew best was a patient and are excluded from the present analysis.

(b) In 1989, two persons with AIDS whose relation to the respondent was that of "patient" are excluded from the tabulations.

(c) In 1990, 17 persons with AIDS whose relation to the respondent was that of "patient" are excluded from the tabulations.

(d) The CDC categories are White, not Hispanic; Black, not Hispanic; Hispanic; and Other/Unknown (Other includes Asian/Pacific Islander and American Indian/Alaskan Native). The GSS question asked "What (is/was) that person's race? (Is/Was) it black, white, hispanic, or other?"

(e) In the GSS in 1988 respondents were only asked specific questions about the person they knew best. Additional persons with AIDS characteristics for race/ethnicity were calculated from a crosstabulation of the total number of people with AIDS known to respondent, the race/ethnicity of the closest victim, respondent's race, and whether respondent is Hispanic or not (primary national origin of Mexican, Puerto Rican, or other Spanish).

(f) Age distribution for data from CDC is actually: Less than 13, 13-19, 20-39, and 40 and above.

(g) Region data for 1988 comes from the CDC report of May 16, 1988. Region data for 1989 comes from the CDC monthly report of August, 1989. The Other category which accounts for about 2% of the total cases is made up of Puerto Rico, the Virgin Islands, Guam, and the Trust Territory.

Figure 1

Respondents Who Know A Person With AIDS, August 1987 - March 1990

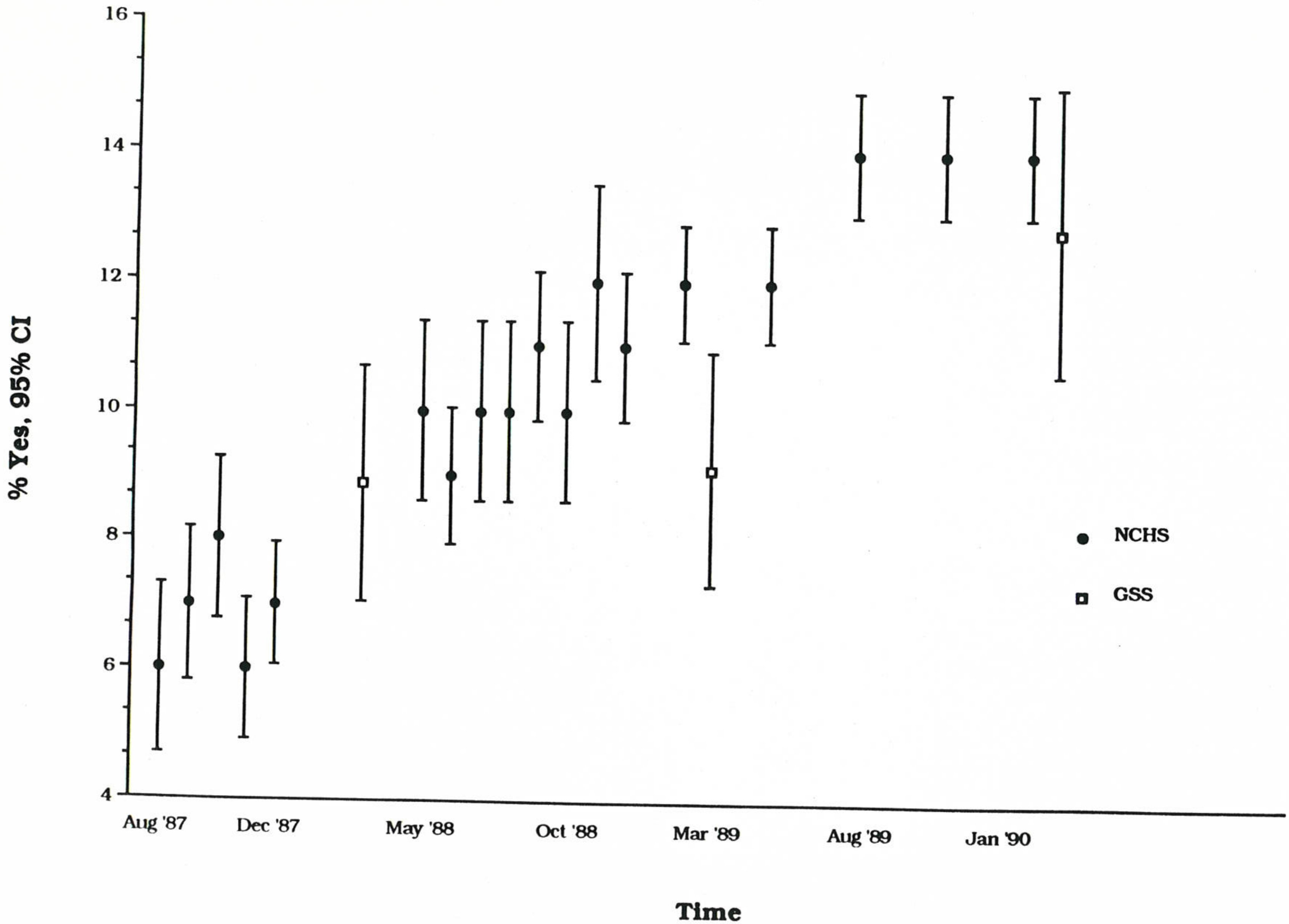
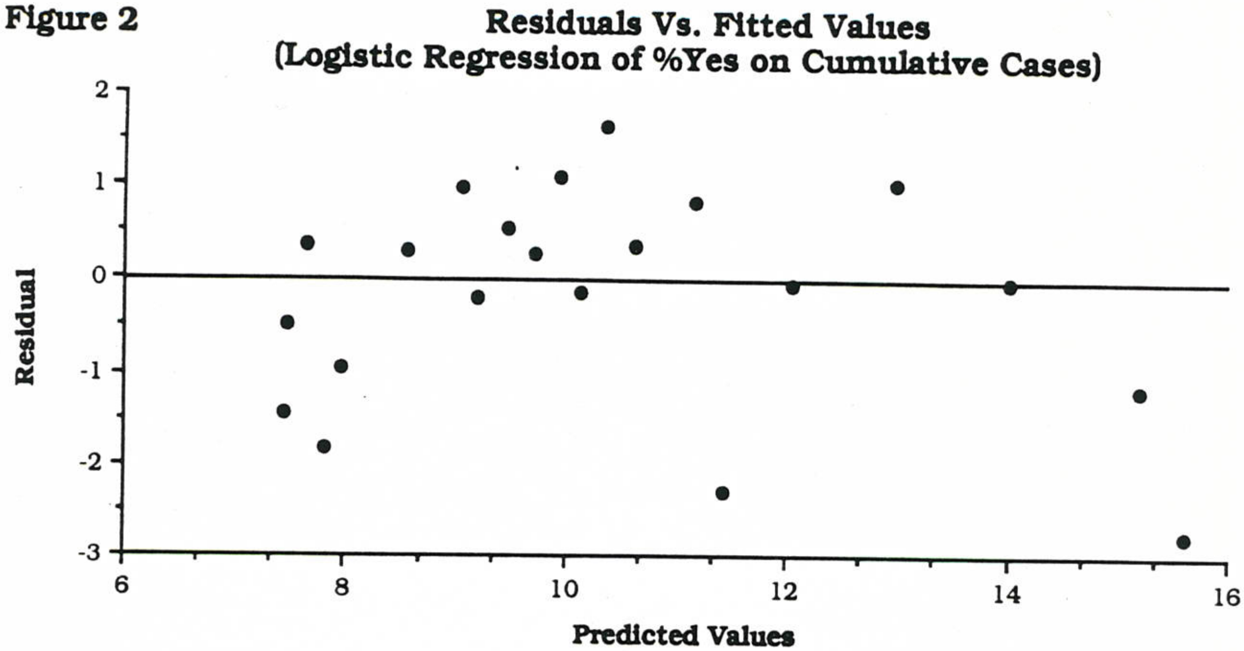


Figure 2


Residuals Vs. Fitted Values
(Logistic Regression of %Yes on Cumulative Cases)
Residual
2
1
0
-1
-2
-3
6
8
10
12
14
16
Predicted Values

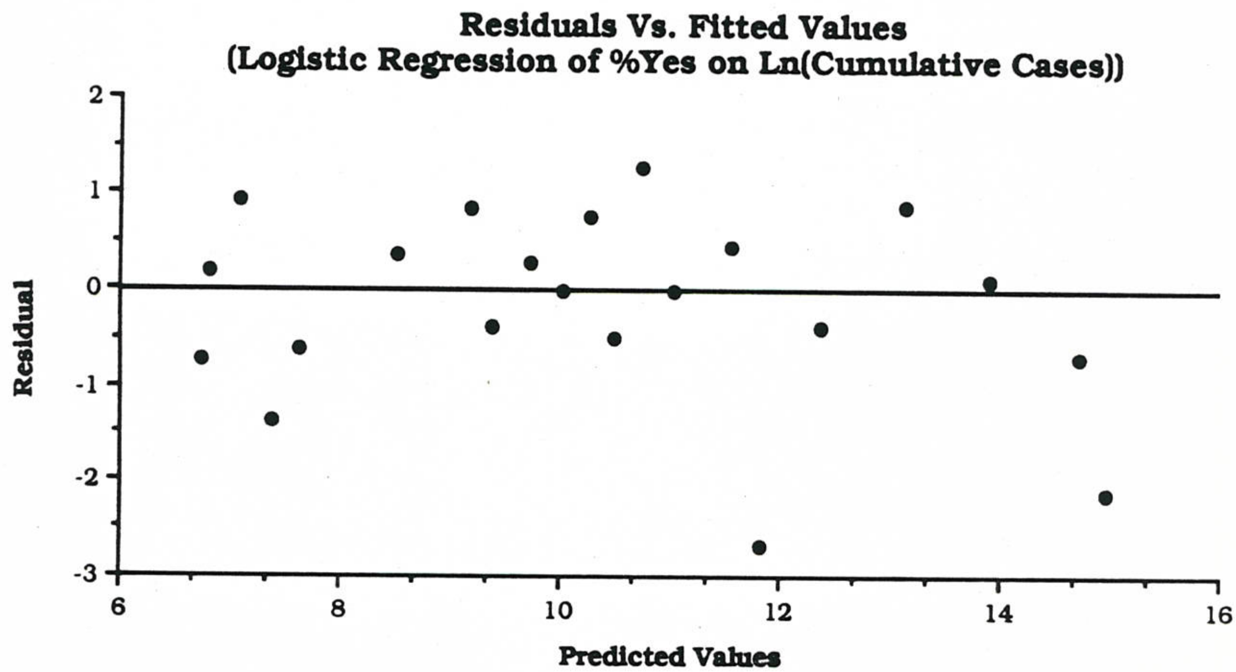
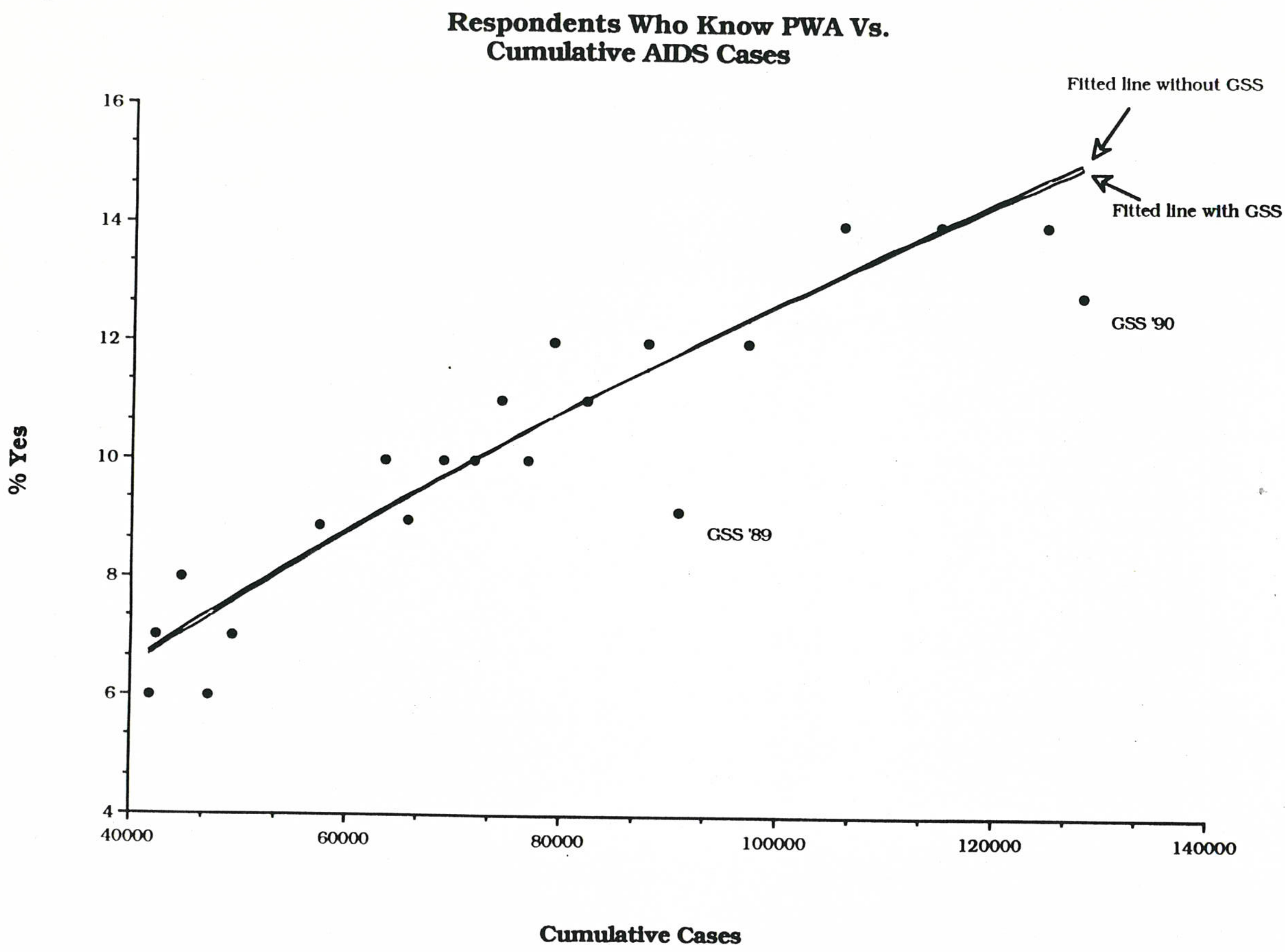

Residuals Vs. Fitted Values
(Logistic Regression of %Yes on Ln(Cumulative Cases))
Residual
2
1
0
-1
-2
-3
6
8
10
12
14
16
Predicted Values

Figure 3



Respondents Who Know PWA Vs.
Cumulative AIDS Cases
% Yes
16
14
12
10
8
6
4
40000
60000
80000
100000
120000
140000
Cumulative Cases
GSS '89
GSS '90
Fitted line without GSS
Fitted line with GSS

Table 4. Logistic Regression Models

	Including GSS			
	Intercept	Slope	t-Statistic	Chi-Square(1)
Age:				
18-29	-3.064	0.3518	11.7	154.8
30-49	-2.627	0.3158	15.4	264.2
50+	-3.486	0.3567	12.2	170.7
Education:				
LT 12	-3.639	0.3021	5.8	36.5
12	-3.539	0.4332	12.8	180.9
MT 12	-3.327	0.5966	22.0	558.4
White	-3.066	0.3547	21.7	535.5
Black	-2.535	0.2682	7.8	67.3
All	-2.997	0.3385	23.2	609.3