Sexual Histories: Does it matter who is asking?

Deven Thomas Hamilton

University of Washington

Department of Sociology

2003

Abstract

Sexual Histories: Does it matter who is asking?

Deven Thomas Hamilton

Objective: To compare basic data on sexual behavior gathered in nine different surveys, all of which used nationally representative samples. **Methods**: We compare estimates of sex partnerships in the last year and over the lifetime from nine data sets. Permutation tests are used to assess whether studies have significantly different estimates controlling for the race, age, sex and marital status composition of the sample. Year of data collection and data collection method were also investigated as potential contributors to the differences in reporting. **Results**: Studies showed significant differences at both the bivariate level, and after adjustment for the demographic attributes of their samples. The bivariate effects of the study ID are, in some cases, 2-3 times larger then the main effect controlling for demographic characteristics. In other instances the opposite occurs as the effects of demographic characteristics have a suppressive effect on study ID. In either case study ID was found to be significant, however the unique effect of the study on responses does not appear to be as large as the effect of age, gender, or marital status. Demographic interaction effects are small, and do not diminish the main effects of study. Further, top-coding responses (as many instruments do) increases the size of the "study effect". The year of data collection was not found to contribute significantly to differences is reporting of sexual behavior. The effects of data collection method were found to be significant with face to face interviews generating higher estimates of sexual partners than either telephone surveys or self-administered questions. This trend held for both males and females. **Conclusion**: There are significant differences in responses to similar questions regarding sexual behavior across the nine nationally representative surveys examined here.

TABLE OF CONTENTS

List of Figures ……..…….…..ii

List of Tables …….…....iii

Introduction ....……..…...1

Data …...……...…10
Surveys ……...……...…10
Adjusting the data and measures……..…..…14

Study design ……...…….…..25

Methods …… ……..…..27

Results ……...……30

Discussion ……..…34

Conclusion ……..…....…..40

Bibliography ……... ……..……...…42

Appendix A: Figures …..……..…..…..46

Appendix B: Tables ……..……...…55

LIST OF FIGURES

Figure 1. Line graphs and error bars for estimates of sex partners over the lifetime and in the last year reported as a continuous variable……....46

Figure 2. Line graphs and error bars for estimates of sex partners over the lifetime and in the last year reported as a top-coded variable……47

Figure 3. Box-plots of the f-statistics generated from the permutation tests of partners in the lifetime and the last year as a continuous variable……...….48

Figure 4. Box-plots of the f-statistics generated from the permutation tests of partners in the lifetime and the last year as a top-coded variable……..……..….49

Figure 5. Box-plot of the f-statistics generated from the permutation test of partners in the lifetime as a top-coded variable for the adolescent subgroup…….……50

Figure 6. Estimated number of lifetime partners as a continuous variable from surveys utilizing SAQ and surveys utilizing FTFI plotted by demographic group…….……51

Figure 7. The mean estimates of lifetime partners reported in surveys using SAQs and surveys using FTFI for females and males…….…..52

Figure 8. Box-plots of the f-statistics generated from permutation tests of surveys which used FTFI to get estimates of partners in the lifetime and partners in the last year……53

Figure 9 Box-plots of the f-statistics generated from permutation tests of a survey which used FTFI and a survey witch used SAQ to get estimates of partners since age eighteen……..…...54

LIST OF TABLES

Table 1. An overview of the relevant attributes of the nine population-based surveys used in the study……55

Table 2. The questions used in each of the nine surveys and a brief description of the adjustments made to the data……...56

Introduction

With the rapid emergence of incurable life threatening STIs like HIV, sex behavior data has become increasingly more important for public health research and prevention planning. Little is known about the reliability of such data, but expectations tend to be that sensitive data often have low reliability (Weinhardt et al. 1998). Further, it has been suggested that self reports of sexual behavior are inherently unreliable and invalid (Lewontin 1995) particularly when the behavior in question confers risk for HIV (Brody 1995). Sexual behavior data is also particularly difficult to assess because the private nature of the subject is not conducive to alternative measurement techniques such as direct observation, nor do the behaviors in question leave secondary evidence such as cigarette butts or candy bar wrappers. Thus, there is presently no "gold standard" that can be used to evaluate the accuracy of estimates for either lifetime sexual partners or annual sexual partners.

However, it is my opinion that much of the previous research has been shown in an overly pessimistic light. While there are likely to be some differences in reporting across surveys, it is my contention that reporting of partners in the last year and over the lifetime will be more consistent across nationally representative surveys then the previously stated opinions of Weinhardt et al, Lewontin and Brody would suggest.

The "number of partners" respondents report is the focus of this analysis because of the central role partner number plays in the study of STIs including HIV. Research on treatable STI transmission dynamics has focused on "core groups" of repeaters. The core

group theory suggests that a small subset (core group) of the population which is characterized by large numbers of sexual partners and repeated infection contribute disproportionately to both the treatment caseload and the spread of STIs (Hethcote and Yorke1984; Thomas and Tucker 1996). Mathematical models suggest this small subset may act as a reservoir which perpetuates endemic transmission even when contact rates are below the epidemic threshold in the general population.

One of the major contributing factors to the rate of infection is the number of sexual contacts between infectious individuals and susceptible individuals which is, in turn, a function of the mean number of new partners each individual in the population has per unit time (Hethcote and Yorke 1984; Anderson and May 1992). Thus the effective use of core group theory depends on accurate accounts of sexual partnerships.

While the core group approach has been, and continues to be, an effective tool for modeling and understanding infectious disease dynamics, HIV and other incurable STIs do not fit the core group model because there are no repeat cases. The recent and rapid proliferation of incurable STIs, like HIV, which do not fit more tradition models of transmission and spread, have led to a shift in the way epidemiologists approach the question of transmission dynamics, now placing more emphasis on the analysis of social networks. Focusing attention on broader partnership networks rather then individual behavior has provided new insight into the spread HIV/AIDS and other STIs. Klovdahl (2000) provides a detailed review. However, network approaches to the study of STIs requires reliable information on contacts between individuals. Just as the core group approach required the mean number of sex partners in order to make accurate

calculations, network analysis requires accurate information on the ties between persons in the network.

There has been a great deal of research done to investigate the reliability of self-reports of sexual information using test-retest methods. These studies have focused on a wide range of behaviors and population including gay men (Saltzman et al.1987; McLaws et al. 1990), heterosexual men and women (Durant and Carey 2002; Van Duynhoven et al. 1999; Weinhardt et al. 1998b; Taylor et al. 1994), different racial groups (Sneed et al. 2001; Kalichmann et al. 1997), and the mentally ill (Carey et al. 2001; Sohler et al. 2000). These studies have generated test-retest correlations from .3 to .9 (Many of these studies are reviewed in Cantania et al. 1995). Given the range of correlations, Durant and Carey have raised the two part question "Is the assessment of sexual behavior uniquely difficult to report, or are stable estimates of most health related behavior difficult to elicit (Durant and Carey 2002)?" In their analysis, they find that both face to face interviews and self-administered questionnaires yield reliable reports of both sexual and non-sexual behavior when the time periods under consideration are extremely short (Durant and Carey 2002).

Other research in this general area has also yielded ample evidence indicating that embarrassing or socially undesirable behaviors are often systematically misreported in surveys. Broad reviews of this literature are provided by Bradburn (1983), Catania et al. (1990) and Weinhardt et al (1998a). Such misreporting may be the result of non-response or non-disclosure. Eligible persons may refuse to participate in the survey, or they may decline to answer specific questions. In both cases, the respondents with the most sensitive information to report may be the least likely to report it (Tourangeau and

Smith 1996). Another source of reporting error is systematic misreporting; the mere fact that a respondent provides an answer is no assurance that the information provided is accurate.

Some evidence suggests that nondisclosure can be reduced through modifications in survey design. If respondents are reluctant to admit that they have engaged in illegal or otherwise embarrassing activities, they may be more forthcoming when they do not have to give information directly to an interviewer. This hypothesis is supported by several studies which have demonstrated that self-administration of sensitive questions increases levels of reporting relative to administration of the same questions by an interviewer. Tourangeau and Smith (1996) provide an extensive listing of such studies. More specifically, it has been shown across a range of items involving sexual behavior and drug use that Audio Computer Assisted Self-Interviewing (ACASI) and Computer Assisted Self-Interviewing (CASI) generally yield higher levels of reporting then does Computer Assisted Personal Interviews (CAPI) (Tourangeau and Smith 1996).

Self-image may also systematically distort reporting across different demographic groups. For example, men consistently report more opposite sex partners than do women, a difference which persists even when differences in population sizes are taken into account (Smith 1992). However, it has also been suggested that these discrepancies in reporting across sex only manifest in the extreme tails of the distribution and may not affect the majority of the data (Morris 1993). Both of these findings are consistent with the view that responses to questions about sexual behavior are strongly affected by self-image concerns as is suggested by (Tourangeau and Smith 1996).

Another potential source of systemic reporting error is questionnaire design. In surveys, prior items have been shown to offset responses to later items and the format of responses can affect the level of reporting. For example, women tend to report larger numbers of sexual partners when the available response categories are geared toward larger responses (Smith 1992; Tourangeau and Rasinski 1988). It has also been shown that respondents systematically eliminate information from their responses if the information has already been collected through previous questions (Schwarz et al. 1991; Tourangeau et al. 1991). For example, if a respondent has been asked about partners in the last year, he or she may eliminate those partners in answering questions about partners in the lifetime. Respondents also tend to heap estimates around common numbers such as five or ten, a process which becomes even more prevalent as the estimates grow larger (Huttenlocher et al. 1990). For a more complete review of the effects of survey and question design see Schaeffer and Presser (2003).

General awareness of these potential problems, while both important and increasing, has not led to a convergence of opinions on the extent of misreporting or the relative impact of the various underlying mechanism. Regardless of the variegated opinions on the specific causes of misreporting there is a persistent expectation that potential respondents are reluctant to divulge sensitive information about their sexual behavior. However response rates tend to be fairly high, particularly when juxtaposed with historically low response questions like income, so non-response may be less of a problem then intuition might indicate.

Currently there are at least nine population based surveys available that provide data on the number of sexual partners a person has had in the last year and/or over their

lifetime. These surveys are not identical in purpose, design or implementation, but they are sufficiently similar that a comparison of their results could shed light on the consistency of sexual behavior reporting in the survey context. The purpose of this paper is to examine the consistency of partnership data using two measures of sexual behavior: The number of partners respondents have had in the last year, and the number of sexual partners the respondents have had in their lifetime. Without a "gold standard," I am not in a position to say anything definitive about the validity of these data, but rather, I am trying to show to what extent these data are consistent. That is to say, to what extent do responses differ across survey when the differences in the surveys themselves are taken into account. To my knowledge no study to date has analyzed responses to sexual behavior questions across such a range of nationally representative surveys.

The nine surveys I include in this analysis are the Behavioral Risk Factor Survey (BRFS), General Social Survey (GSS), National College Health Risk Behavior Survey (NCHRBS), National Health and Social Life Survey (NHSLS), National Survey of Family Growth (NSFG), National Survey of Men (NSM), National Survey of Women (NSW), Youth Risk Behavior Survey (YRBS), and the National Longitudinal Study of Adolescent Health (Add Health).

All of the samples utilized in these surveys are population based, but they are not all probability samples. Thus, some variation might be expected to arise from differences in the sample. Further, the differences in the instruments, either the mode of administration or the construction of the questions themselves, are sufficient to allow for the possibility that they are all reliable and yet not consistent. Reliability is "a matter of whether a particular technique, applied repeatedly to the same object, would yield the

same result each time" (Babbie 1995). For this paper it is true that the same technique, "a survey", is being applied to the same object, "a nationally representative sample." However, the survey methods and sample do vary, thus reliability can not be directly assessed. Consistency can show the degree to which deferent techniques applied to different objects generate the same results. Thus, while consistency and reliability are similar in that they are measures of the degree to which results are the same, they are not necessarily linked. Consistency does not connote reliability, and a lack of consistency is not indicative of unreliability. Comparisons of these studies in terms of consistency can tell us something about the sources of differentiation and bias, as well as provide us with a more general understanding of the degree to which survey results may differ as a general function of the survey, the specific attributes of particular surveys, or differences in sample composition.

There has been some previous work evaluating the validity and reliability of these surveys as sources of sensitive material. Abortion rates reported as part of the NSFG have been compared to abortion rate data collected from abortion clinics. The NSFG yielded estimates far lower than those gathered from the clinics suggesting that fewer than half of all abortions are reported in the NSFG (Jones and Forrest 1992). The low level of reporting for abortions suggests that there may have been a general tendency towards under reporting when questions administered as part of the NSFG addressed sensitive issues. Santelli et al. (2000) compared trends in sexual behavior among adolescents overtime using data from the NSFG, YRBS, National Survey of Adolescent Males and Add Health. They identified significant trends in some studies which were not

borne out in others suggesting that studies do generate inconsistent findings on sexual behavior.

In this analysis, I investigate the consistency of data on the number of sexual partners reported by respondents in these surveys, while controlling for basic demographic characteristics. It is not my intension to make any broad claims about the reliability or validity of these data, but rather to determine the level of consistency across surveys and identify some of the specific differences in surveys which generate differences in reporting.

My primary analysis focuses on the general differences across surveys as a general function of the survey. In this context the survey is a collection of all of the differences not accounted for by sample composition which I control for. In addition to this exploration of the general differences, I also analyze the effects of several factors that are likely to contribute to any differences across surveys. The first of these is data collection method. As was stated previously, there is evidence to suggest that self-administered questionnaires (SAQ) will increase the level of reporting of sensitive information relative to face to face interviews (FTFI) (Tourangeau and Smith 1996). The surveys analyzed here used three data collection techniques: telephone interview, SAQ and FTFI. The analysis of data collection method only utilized those surveys that included adult populations. Adolescents were not included in the analysis of data collection method because only two surveys included adolescents. As such the data collection method and study ID are perfectly correlated; making it is impossible to separate study effect and the effect of data collection.

The second potential source of variation I investigate is the year of data collection. Two of the surveys I use are repeated surveys, so it is possible to investigate changes over time while controlling for the survey method, design and content. I first look at the independent effect of year, that is, do responses vary significantly by year. I then investigate whether there are trends in reporting by year, that is, are reports increasing or decreasing over time.

I also investigate the effect of different measurement strategies, specifically the effect of top-coded response categories verses open-ended responses.

Data

There are many data sets currently available which include questions about sexual behavior. I focus on 9 population based studies, the BRFS, GSS, NCHRBS, NHSLS, NSFG, NSM, NSW, YRBS, and Add Health (table 1 provides a general summery of the data sets). There are several differences between these surveys which require making several adjustments to the data to facilitate comparability. Here I include a description of these data sets along with some explanation of the adjustments made to make them comparable.

Surveys:

The BRFS is a part of the state-based Behavioral Risk Factor Surveillance System initiated in 1984 by the Center for Disease Control (CDC) to collect prevalence data on risk behaviors and preventative health practices. The questionnaire was developed jointly by the Behavioral Surveillance Branch of the CDC and the individual participating states. The survey was conducted via telephone using telephone numbers drawn from two strata based on the presumed density of known household numbers. Any adult eighteen years of age or older was eligible to participate in the survey (Centers for Disease Control and Prevention 1998).

The GSS was designed as part of a program of social indicator research which focuses on replication of questionnaire items and wording in order to facilitate time-trend studies. The sample for the survey was the National Opinion Research Center national probability sample which includes all non-institutionalized English-speaking persons

eighteen years of age or older, living in the United states. Respondents reported estimates of sexual partners via SAQ. The GSS samples were designed to give each household an equal probability of inclusion in the sample, and the sample distribution closely resembles the distributions reported in the Census and other authoritative sources. However, due to non-response, bias, sampling variation etc., the sample does deviate to some degree from the actual population (The National Opinion Research Center 2003).

The NCHRBS is a section of the Youth Risk Behavior Surveillance Survey developed in 1990 by the CDC to monitor health-risk behavior among young persons. The NCHRBS utilized a mailed SAQ, was performed in 1995, and was specifically designed to assess health-risk behaviors among college students. The NCHRBS used a two stage cluster sample design to produce a nationally representative sample of undergraduate students aged greater than or equal to eighteen. The first-stage-sampling frame contained 2,919 primary sampling units, consisting of two and four year colleges and universities. Of these 148 were selected from sixteen strata formed on the basis of the relative percentage of black and Hispanic students. The second-sampling-stage consisted of a simple random sample draw from a list of the full and part time undergraduate students aged eighteen or greater enrolled in the 136 participating colleges and universities. Of those eligible, 4,838 completed the questionnaire (MMWR Nov. 14 1997).

The NHSLS is a 1992 survey of non-institutionalized 18 to 59 year-olds living in the United States who where able to complete the interview in English. The study used FTFI to collect data on sexual experiences and other social, demographic, attitudinal and

health related characteristics. Respondents were selected using a multistage area probability sample designed to give each household an equal probability of inclusion. A cross sectional sample of 3,159 respondents was collected as well as and an over-sample of 273 black and Hispanic respondents (Laumann et. al. 1995).

The NSFG is a multipurpose survey of a national sample of non-institutionalized women 15-44 years-old residing in the United States which was sponsored by the National Center for Health Statistics. This analysis utilizes data from cycle V of the survey which was carried out in 1995 using FTFI. The sample for cycle V is a national probability sample of 10,847 women aged 15-44 from household that had participated in the National Health Interview Survey (1993). The survey over-sampled both black and Hispanic women (U.S. Dept. of Health and Human Services, National Center for Health Statistics 2000).

The NSM 1991, was designed to examine issues related to sexual behavior and condom use. The study population consisted of 20-39 year old non-institutionalized men. The sample was based on a multi-staged, stratified, clustered, disproportionate-area probability sample of households within the contiguous United States and included an over-sample of Blacks. Data on sexual partners was collected using FTFI (Tanfer 1993).

The NSW was also conducted in 1991 and was designed to examine sexual, contraceptive, and fertility behaviors, and the factors associated with these behaviors in a nationally representative probability sample. The survey includes a total of 1669 cases from two sub-samples. The first of these sub-samples (n=929) were follow-up cases from the 1983 National Survey of Unmarried Women which sampled 1314 never-married

women between 20 and 29 years of age. The second sub-sample (n=740) is from a second probability sample of 20 to 27 year old women of unspecified marital status which was conducted in 1991. Data on sexual partners were acquired using FTFI (Sociometrics Corporation 1990 and Tanfer 1995)

The YRBS was conducted in 1992 as a follow back and supplement to the National Health Interview Survey, which was initially designed to obtain information about the amount and distribution of illness, its effects, and the kinds of health services people receive. Specifically, the YRBS, which was sponsored by the Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, was developed to monitor the major risk behaviors of American youth. Respondents were civilian residents of the United States who were non-institutionalized and aged 12-21. Respondents were randomly sampled from each household in the 1992 NHIS sample. Within each family, one child who was attending school, and up to two children not in school or "status unknown" were selected for the survey (US Department of Health and Human Services 1992).

The Add Health survey, which was funded by the National Institute of Child Health and Human Development and seventeen other federal agencies, was conducted from September 1994 through December 1995, and then again from April 1996 through August 1996. It was designed to investigate health-related behaviors and there causes among adolescents. The study was school-based and focuses on adolescents in seventh through twelfth grade. The survey consisted of a national sample of schools in the United States and included both in-school surveys and in–home interviews. 90,118 students

completed the original in-school questionnaires between September 1994-April 1995, while 20,745 adolescents completed the in-home interviews for Wave I and 14,738 adolescents completed the in-home interview in wave II. (The National Longitudinal Study of Adolescent Health 1998)

Adjusting the data and measures:

There are obviously considerable differences between these surveys which require rectification before the surveys can be compared in a meaningful way. Some of these differences are explicit differences in the design and or implementation of the survey while other differences arise from the different approaches to measurement of estimates.

Examples of differences in the design of the surveys include different time horizons for sexual partnerships of respondents, the sex of eligible respondents and the age of eligible respondents. Some of the differences in measurement include the phrasing of questions, or whether the number of partners in a given time period is ascertained from a single question or multiple questions. For example, the NSM does not ask about partners in the last year or lifetime but rather asks specifically about the number of anal sex partners and the number of vaginal sex partners in particular time periods. Similarly, the variable used from the GSS for lifetime number of sex partners is actually only partners since eighteen. As a result the GSS is expected to have slightly lower responses then other surveys. An additional difference which is particularly important for this analysis is the structure of responses. Some studies used continuous variables while others used a top-code or categorical variables.

In some cases it is possible to find a lowest common denominator, for example applying the lowest top-code across all relevant studies. However, some differences proved insurmountable, so I adopt the approach of conducting several separate analyses utilizing different but overlapping combinations of surveys. Here I provide a detailed description of the adjustments I made to each of the data sets (Table 2 provides a general overview of the major adjustments). While it may be difficult for readers to keep track of all of the individual adjustments due to the number of surveys involved, it is sufficient to bare in mind a simple rubric which I used to inform all of the decisions regarding data adjustment and variable construction. For all surveys I sought to capture the maximum number of *distinct* partners reported by respondents.

The BRFS, which provided data across the five years 1996 through 2000, only included a question about partners in the last year. As such, it is only included when partners in the last year is the dependent variable of interest. In the survey the questions regarding sexual behavior are part of a supplement which states can elect to include each year. Thus, not all states are included every year. Additionally, all questions regarding sexual behavior were only asked of respondents less then 50 years of age. Therefore, all respondents over the age of 50 were eliminated. Additionally, the variable for the number of sex partners in the last year is top coded at 76+. There are three respondents reporting 76+. As a result this survey is only compared to other surveys when "sex partners in the last year" is top-coded to 76+. From 1996 thru 2000, a total of 72,280 respondents reported the number of sex partners they had had during the previous year.

The GSS asked questions about the number of sexual partners in the last year and over the lifetime in 1989-1991, 1993, 1994, 1996 and 2000. Additional data on the number of partners the respondent had in the last year was also reported in 1988. There are 16,159 responses to the number of sexual partners in the last year and 14,847 responses to the number of sexual partners over the lifetime. Unfortunately the GSS did not specifically ask about lifetime partners so I construct a proxy from two closely related questions: male partners since eighteen and female partners since eighteen. As a result the GSS is expected to have slightly lower responses then other surveys. When the estimates for partners in the last year and lifetime are compared, 697 cases report more sex partners in the last year then in their lifetime (since eighteen).

There are two potential sources for this discrepancy. The first source is the wording and location of the questions. The question regarding partners in the last twelve month preceded questions about partners since eighteen which may have led respondents to report the number of sex partners since age eighteen excluding the partners in the last year. Schwarz et al. 1991 and Tourangeau et al. 1991 report that respondents systematically eliminate information from their responses if the information has already been collected through previous questions. Of the 697 (4% of the total number of respondents) respondents reporting more sex partners in the last year then in their lifetime, 522 report only one partner in the last year and zero partners in their lifetime. 73 of them report two to four partners in the last year and zero in their lifetime. The pattern of responses for these 595 respondents is consistent with aforementioned alternate interpretation of the questions. The second source could be the result of an overlap in the

start and end dates of last year and since eighteen. If the respondent is close to eighteen and had more partners in the last year then since turning eighteen the number of partners in the last year could well be greater than the number of partners since turning eighteen. Only three respondents who are less then or equal to eighteen report more partners in the last year than they report having since turning eighteen. Thus it seems that the vast majority of the discrepancy between partners in the lat year and partners in the lifetime is likely the result of a misinterpretation of the question rather than a result of the respondent's age. In order to deal with this discrepancy, lifetime partners are recoded to be the sum of lifetime partners and partners in the last year if and only if the number of partners in the last year was greater then lifetime partners.

The GSS variable for the number of sex partners in the last year is categorical and top-coded (1, 2, 3, 4, 5-10, 11-21, 21-100, 100+). As a result, these data are not directly comparable to the data collected in several of the other studies. In order to include the GSS data for partners in the last year, responses in other studies are top-coded to 5+ when they are analyzed along with the GSS data. The GSS is not included when partners in the last year are analyzed as a continuous variable.

The NCHRBS did not ask a single question about the number of partners in the lifetime, so I construct the variable from two questions: the number of male sexual partners and the number of female sexual partners over the lifetime. Each of these questions was top coded at 6. In order to get an estimate of the total number of sexual partners, the responses to these two questions are combined but the top-code remains in place allowing for a maximum of six partners. These data are only compared to other

surveys when a top-code of six is either present in the study design of the other studies in the comparison or when a top-code is imposed on the other studies by the researcher. The NCHRBS did not include questions on partners in the last year so it is not included in those analyses in which partners in the last years are the dependent variable of interest.

I also constructed the variable for the number of lifetime partners for the NHSLS respondents. This variable is constructed from 14 different variables reported by the NHSLS. There are four questions on the NHSLS which ask about sexual partners before age 18. One of these asks about the first opposite sex partner, another about the first same sex partner. The other two ask about any other opposite sex partners or same sex partner before eighteen respectively. Thus, the number of partners before eighteen is constructed by adding the number of other same sex partners before 18 to the number of other opposite sex partners before 18. Then an additional partner is added for the reports of first same sex partner and first opposite sex partner if they occurred before the respondent was 18. In order to determine the number of partners since eighteen I follow the strategy outlined by Laumann et al. 1994 in appendix 5.2A. However I do not include data from the SAQ because the NHSLS is treated as a FTFI based survey in these analyses.

In the NHSLS the number of partners in the last year is represented by several variables. The first variable is given as a set of raw numbers, while the second variable is a constructed variable which was generated by the original researchers using multiple reference points throughout the survey process. Laumann, et al. argue that the constructed measure is more accurate because it was constructed using various reference

points during the survey. However, the constructed variable is categorical and the data used for the constructed variable are drawn from both face to face interviews and self-administered questionnaires. I am particularly interested in continuous responses, as well as the effect of the data collection method. I therefore only use the continuous variable which was asked in the face to face interview.

The NSFG provides data on both the number of sexual partners in the last year and the number of sexual partners over the respondent's lifetime. For both questions the instrument used by the NSFG allows respondents to give estimates in the form of a high and a low boundary if they can not recall an exact number. Only 40 respondents (.4%) report a high estimate of the number of partners they had had in the last year that is different from the low estimate. For the number of partners in the lifetime, 468 (4.3%) respondents report a high estimate that is different than the low estimate. Of those 468 respondents, 176 of them report a high estimate that is only one higher then the low estimate. In order to reduce responses to a single variable I construct a new variable by taking the mean from the high/low estimates. To adjust for resultant half partners, responses are rounded to the nearest whole number. There are 2 cases reporting more sex partners in the last year then in their lifetime. To adjust for this discrepancy lifetime partners is set equal to partners in the last year if partners in the last year are reported to be greater then partners in the lifetime.

The instrument used in the NSM does not directly ask respondents about the total number of sex partners respondents have had in the last year or over the life course. Rather, respondents are asked to provide the number of vaginal sex partners they have

had in each of the time periods of interest and the number of anal sex partners they have had in each of the time periods of interest. Unfortunately there is no way to ascertain how many partners could have been represented in both categories. Therefore, I estimate the number of partners to be equal to the variable for which the respondent provides the highest response. As a result of this adjustment I expect that the estimates I use for the NSW are lower then what the respondents actually reported. It should be noted that 20% of the respondents report having had anal sex and 3% of respondents report more then 5 anal sex partners. However, only 35 respondents report having had more anal sex partners then vaginal sex partners, and eighteen of those 35 report never having had a vaginal sex partner. Finally, if the number of partners reported in the last year is greater than the number of partners reported for the lifetime, lifetime partners is set equal to the last year. There are only 5 such cases.

As was the case with the NSM, the instrument used in the NSW asks about the number of partners with whom respondents had vaginal sex and about the number of partners with whom they had had anal sex. Again there is no way to differentiate between partners that fit into only one of those categories and partners that are in both categories. Therefore, if respondents reported both vaginal sex partners and anal sex partners I used the same strategy as was previously mentioned for the NSM. Only 216 of the women surveyed report having had anal sex, and of those only 18 reported doing so with more than 2 partners. As a final adjustment, if the number of partners reported in the last year is greater then the number of partners reported for lifetime, lifetime is set equal to the last year. There are only 14 such cases.

The YRBS provides data exclusively on the total number of lifetime sex partners. The responses to questions regarding lifetime sex partners were top-coded at 6+ which may result in low estimates for some respondents. These estimates are only compared to estimates from Add Health, the other survey of adolescents.

The Add Health survey included a question in wave I of the survey which asked about the number of sexual partners respondents had had in their lifetime. This value is added to a second variable in wave II which asked for the number of partners since the last interview. I top-code the variable so it is comparable to the YRBS.

Each of the data sets used for these analyses had unique implementation, sources of bias and difficulties. Many of the idiosyncrasies in each data set are in part accounted for by the weights provided by the original researchers who collected the data. Several of the weights include adjustments for demographics that I account for through the inclusion of demographic attributes as covariates in my analysis. However, the different sampling strategies and rates of non-response which are also adjusted for by the weights provided with the data sets are not accounted for by covariates. Therefore I weight the different data sets by the weights provided by the original researchers.

The data from the BRFSS is weighted with the variable "finalwt" which accounts for the probability of reaching a household, the probability of randomly choosing an adult from a household, and the demographic distribution of the sample. The final weight for the BRFS adjusts the data to match the distribution of the actual population as well as the size of the population it samples (Centers for Disease Control and Prevention 1998). For the GSS, which does not provide a post-stratification weight, I apply a weight for the

probability of selection in to sample from each household. Each interview is simply weighted proportionally to the number of eligible respondents in each household (The National Opinion Research Center 2003). The NCHRBS provides a final weight which includes adjustments for the school sampling weight, the school non-response adjustment factor, the student sampling weight, the student non-response adjustment factor, and a post-stratification to account for the total number of students by race/ethnicity, sex, and institution type (MMWR Nov. 14 1997). A final weight is provided in the NHSLS that adjust for the different probabilities of inclusion for the over-sample and cross-section cases while also adjusting for household size and for non-response (Laumann et. al. 1995). For the data from the NSFG I apply the weights provided which match the data to the nationally representative estimates that are reported by the National Center for Health Statistics (U.S. Dept. of Health and Human Services, National Center for Health Statistics 2000). The NSM is weighted using the final weights provided with the data which includes a sampling weight, screening weight, eligibility weight, non-response weight and post-stratification weight (Tanfer 1993). The data from the NSW is weighted using a final sample weight which is a composite based on two intermediate sample weights from the new women's sub-sample and the 1983 re-interview sub-sample . Sample weights account for stratification and clustering, differential selection probabilities, over-sampling and non-response (Tanfer 1993). The data from the YRBS is weighted to adjust for non-response, the probability of selection and the over-sample. The final weights are also scaled to keep the sample size constant and the proportion of students in each grade

consistent with the national population (US Department of Health and Human Services 1992).

The demographic variables are recoded into the following categories for comparisons across surveys of adults which include the BRFS, GSS, NCHRBS, NHSLS, NSFG, NSM and NSW. Race is collapsed to three categories: White, Black and Other. Marital status is collapsed into three categories: Married, Divorced-Widowed-Separated, and Never Married. Age is also reduced to three categories: 18-24, 25-34 and 35-44. In some cases, age could not be perfectly collapsed and as a result some of the surveys do not provide data on individuals of every age within each bin.

For comparisons across surveys of adolescents, which include the YRBS and Add Health, the demographic variables are recoded into the following categories. Race is collapsed to three categories: White, Black and Other. Marital status is not included because marital status is not a question in the YRBS. It is unlikely that there would have been enough married respondents to include the variable even if it had been asked due to the relatively young age of the respondents. Age is reduced to four categories: 14, 15, 16 and 17. Respondents between eleven and thirteen were not included because there were so few of them. Older respondents, those eighteen and older, are excluded because the YRBS used a top-code for age and I believe that the respondents in the YRBS reporting ages greater then or equal to eighteen are not similarly distributed to the respondents in the Add Health survey who report age greater then or equal to eighteen.

These variables along with sex are indexed to create 54 unique demographic categories for the adult data and 24 for the adolescents. Using these demographic

characteristics allows me to account for some of the differences in the sample composition of the various surveys, thus allowing me to focus on the differences across surveys rather then between demographic groups. Collapsing the demographics categories into only 54/24 unique groups also increased the number of respondents in each unique group which was necessary to reduce the prevalence of low cell counts or zero count cells.

Study Design

Variations in study design make it necessary to conduct the comparisons for subsets of similar studies. Three factors define the subsets: Outcome measurement scale (continuous/top-coded), sample sex distribution (male only, female only, both sexes), and sample age range (14-17 for adolescents and 18-24, 25-34, 35-44 for adults). This results in 15 separate analyses, each on separate but overlapping subgroups of the data. The first two divisions of the data are based on the dependent variable. I compare the surveys with continuous measures in one analysis, and compare all surveys in another, with the continuous variables recoded to the top-code categories (for partners in the last year the analysis of continuous responses actually uses a top-code of 76) . Some surveys are sex specific, so separate analyses are carried out for males and females in addition to an analysis which includes both sexes. The NSM and NSW are merged to represent a single study and included in the two sex comparison because the timing, structure and population were almost identical.

A separate analysis is performed on the YRBS and Add Health which focus specifically on youth populations. Both of these surveys include males and females so I do not perform separate analyses based on sex category.

For the analysis of survey method, the seven surveys which included adults are divided into three groups based on data collection method, telephone survey, FTFI, and SAQ. Only the BRFS used telephone interviews so the effect of telephone interviews is indistinguishable from the effect of the BRFS more generally. Therefore, telephone

interviews, as a method of data collection, is only included in the descriptive analysis, and not the ANOVA or permutation tests. Unfortunately, due to the differences in questions between surveys and answer format, I am not able to compare reporting of both lifetime partners and partners in the last year as both continuous variables and top-coded variables across all survey methods.

I use data from the BRFS from 1996 through 2000 and data from the GSS from 1988 through 2000 and analyze the two surveys separately for both an independent year effect and for a trend over time. The dramatic decline in the number of data points in each demographic group results in several low *n* or empty cells. To determine if the empty cells affect my results, I run the analysis using the previously described demographic groups and with a more condensed version which categorized race as white/non-white and marital status as married/not-married. There is no meaningful difference in the results.

Methods

Standard exploratory techniques are used to determine what adjustments need to be made to the data in order to make the data comparable. Graphical exploratory techniques are used to get a preliminary indication of the differences in reporting across survey in term of general frequency and distribution.

I use analysis of variance to determine if the study, year, or data collection method have an independent effect on the number of partners identified by respondents, as well as to get an indication of the magnitude of that effect. ANOVA is also used to test the sensitivity of my results to small cell counts and the imposition of top-codes by the researchers. In the case of small cell counts, a dummy variable for cells with five or fewer respondents is included in all of the analyses, and it is found that low cell counts do not meaningfully alter my results. Additionally, a dummy variable is created for studies that are top-coded by design in order to determine if they differed significantly from studies that are not top-coded by design, but rather by me. In this case, as with low cell counts, the effects of the variables of interest do not meaningfully change. Thus, neither variable is included in the final analyses.

In all of these analyses I focused on variance from the mean rather than the median. While the median would have been less affected by outliers, the mean is unique in that it can be scaled up to generate the total level of reporting.

I do not have a balanced design which is generally a requirement for the use of ANOVA, and Type I sums of squares in particular. Despite this I elect not to adjust the

ANOVA to use Type III or Type IV sums of squares. Rather, I use Type I (hierarchical) sums of squares and insert the variable for study last in order to account for differences in the sample composition before estimating the study effect. In the case of the interaction effects model, new variables are created for each interaction so that the variable for study can be put into the hierarchical model after both the main effects and interaction effects of all the independent variables. This allows me to isolate the effect of study which in entirely independent of the other covariates while still using Type I sums of squares.

Further, I use a permutation test to compensate for the unbalanced design and the resultant inflated f-statistics, and to determine the statistical significance of observed differences across surveys. The first step in each of the permutation test is to assign a label from a categorical variable (1 to the number of surveys in the analysis) to each observation, indicating what survey the observation came from. An ANOVA is then run on the data with the categorical variable for survey in the model. The resultant f-statistic for "study" is the observed value. That f-statistic for "study" is then compared to a distribution of 200 f-statistics from 200 ANOVAs run on random permutations of the data.

When the data are permuted, observations within a specific demographic group and from the same survey are held together in a block and the blocks are randomly permuted across the range of the variable for identifying study. Blocking respondents into groups by study of origin and demographic group and holding the observations within the groups constant preserves the within group variance which, in this case, is the unexplained variance at the individual level. By keeping the individual level variation

within the blocks, I am able to isolate the variation that is due only to differences in survey. If I had permuted the individual observations, the within group variance would be conflated with the between group variance making it difficult to differentiate between the two.

The analysis of data collection method and year is carried out in a similar fashion. ANOVA is used to calculate an f-statistic for the difference between cohorts. In both cases, inflated f-statistics which resulted from an unbalanced study design are compensated for with a permutation test wherein the data are randomly permuted across collection method or year but within demographic category. In addition to testing for independent effects for year as a factor, I also test for trends across years using linear regression.

Results

Both error bar plots of the mean number of partners reported for the lifetime (fig. 1) show the GSS to have slightly lower numbers of partners reported on average for both men and women. Also in figure 1, the lowest estimates for partner in the last year coming from the BRFS. In figure 2, which shows top-coded estimates of lifetime partners, we see that males reported far fewer lifetime partners on the GSS and NCHRBS than they did on the NHSLS and NSM. When just the females are considered the top-coded estimates for lifetime partners from the NHSLS, NSW and NCHRBS are all very similar while the GSS and NSFG produce noticeably lower estimates (fig. 2).

In order to reduce the effects of differences in sample composition that might be driving the differences seen in the error bars, and get a more perspicuous indication of the independent effects of survey, I use an hierarchical analysis of variance. The resultant f-statistics for "study" are extremely large, particularly when responses are top-coded. Study does not appear to be as important as age, sex or marital status. All three of these demographic characteristics consistently produced higher f-statistics then study. The only other demographic variable I include in these analyses is race, which has an effect similar in magnitude to study.

To determine whether the large f-statistics are actually significant or an artifact of an unbalanced design, I use a permutation test for each unique sub-group of the data. The results of these tests are shown in Figure 3 through 5 and are unambiguous. This figures show a box plot of the f-statistics generated by the 200 ANOVAs run on the permuted data along with the f-statistic generated from the observed data. Figure 3 shows the

permutation test results for both lifetime partners and partners in the last year as continuous variables. The results are also shown for males and females together as well as separately. Figure 4 shows the results of the permutation tests when responses are top-coded. Figure 5 show the results for the adolescent group. The f-statistics generated from the observed data clearly fall outside of the range of values generated by chance in almost all cases. The observed values, which are marked on the box-plots with an asterisk, are the most extreme outlier for nine of the thirteen groups. Of the remaining four sub-groups, the probability that the observed f-statistic is due to chance is less than .05 for three of them. The only sub-group for which the difference across study is not statistically significant at the .05 level is the number of partners in the last year, males and females included, with the response as a continuous variable. For this sub-group the p-value is .145. That test is shown in the upper right panel of figure 3.

In my analysis of method of administration, surveys that used FTFIs consistently generated higher estimates of number of partners in the last year and over the lifetime when compared to studies that used SAQs or telephone interviews. Figure 6 provides an example of the difference in estimates between SAQs and FTFI. This trend toward higher levels of reporting in FTFI held when variables were continuous or top-coded. Further these differences were persistent among both males and females (fig. 7). There is no analysis comparing estimates of partners in the last year as a continuous variable which includes SAQs and FTFIs because none of the surveys which use SAQs provided such data.

When the general effect of study is analyzed across just those studies which use FTFIs (the NHSLS, NSFG, NSM, and NSW) I find that the effect of study is not significant when both sexes are included (fig. 8), in which case it is the NHSLS and NSM/NSW that are being compared. The results are also non-significant when just the males are included, in which case it is the NHSLS and NSM that are being compared. The analysis of the female respondents, on the other hand, did show significant differences across survey. However, when the females are compared the NSFG, which only includes women, is also included. But when the analysis of the female respondents only includes data from the NHSLS and NSW no significant difference is found. Thus there does not appear to be a difference in reporting between the NHSLS and NSM/NSW but the NSFG does appear to be an outlier of sorts.

In addition to these analyses which focused on just the surveys which used FTFIs, I also perform a comparison between the GSS, which uses an SAQ, and the NHSLS, which uses FTFIs. Both of these surveys provide data on partners since age eighteen which allows me to control for some of the error generated by using proxies for the number of partners in the lifetime. When the two surveys are compared for just partners since age eighteen, the difference between them is not significant (fig.9).

Permutation tests performed with the year of data collection as a factor randomly produces f-statistics much larger then the f-statistic generated by the observed data indicating that there is no significant independent year effect. When year is analyzed for trend using logistic regression only one of the three analyses produce significant results. The BRFS shows a significant upward trend in reporting of partners in the last year

($t=6.431$, $p=1.27e^{-10}$), even when demographic composition was taken into account. The GSS on the other hand did not show a significant trend in the reporting of partners in the last year or lifetime.

The results of the analysis of adolescent respondents from the YRBS and Add Health show a clear difference in reporting across surveys. Most of the difference appears to be driven by estimates on the lower end of the distribution. Nearly fifteen percent more of the respondents reported zero partners in the Add Health survey while the YRBS had higher reports of one to three partners. There appears to be no real difference above four partners. The analysis of variance and permutation test show the difference across study to be significant (fig. 5)

Discussion

Initial exploration of these data using graphical exploratory techniques provides contradictory evidence regarding differences in reporting by group across surveys. The shapes of the frequency distributions are in some cases almost identical, as can be seen in figures 1 and 2. However, the subtle differences that exist between the frequencies reported in each survey seem to have large aggregate effects as can be seen in the error bars also displayed in figures 1 and 2. These graphs only show the frequencies of reported numbers of partners across survey; they do not capture the effect that differences in the composition of the samples may have on general summaries of the data. Thus, while they do indicate a substantial difference across studies they are very rough approximations. A more refined assessment of the degree to which the studies vary shows that the differences are significant for reports of both partners in the last year and partners in the lifetime. Figure 3 through 5 clearly show that there are significant effects of "study" on the level of reporting even when sample composition is accounted for. The question then is, which surveys are different and why.

When looking at the error bars presented in figures 1, which presents the sex behavior variables as continuous, it is clear that the GSS produces lower estimates for the number of partners in the lifetime while the and BRFS generates substantially lower estimates for partners in the last year respectively.

The relatively low estimates from the GSS may in be due, in part, to the operationalization of lifetime partners. The GSS did not specifically ask respondents

about lifetime partners, the question on the survey was the number of sex partners since age eighteen. The elimination of partners before age eighteen may be the driving force behind the lower estimates for the GSS. This hypothesis is supported by the comparison between the GSS and NHSLS using partners since eighteen as the dependent variable. In that analysis no significant difference was found.

The low estimates for partner in the last year, which were generated from the BRFS may be the result of data collection method. This is the only study which used telephone interviews as a method of data collection. It may be that respondents under-report the number of partners they have in this context. However, because the BRFS is the only survey which used telephone interviews I have no way of controlling for the method of data collection and thus, I can not provide any evidence that it is or is not the cause of the low estimates.

When the data are top-coded the GSS remains a low estimate for lifetime partners for both males and females, but the NCHRBS is also a lower estimate but only for males while the NSFG is a lower estimate for females. The lower estimates by males for lifetime partners in the GSS and NCHRBS compared to the NHSLS and NSM may be the result of at least two independent conditions. The first is the difference in the mode of administration, the GSS and NCHRBS are administered as an SAQ while the NHSLS and NSM are FTFIs. However, if this is the case it contradicts previous research which suggests that respondents are more forthcoming on SAQs than in FTFIs. The other possibility is that the lifetime estimates for the GSS are deflated by the elimination of partners before age eighteen as was mentioned previously. However, that does not

explain the low estimates generated by the GSS for partners in the last year shown in figure 2. The number of lifetime partners reported on the NCHRBS may be lower when demographic attributes are not taken into account because this survey focused on college aged respondents who lack the years of experience that respondents in other surveys have. Interestingly, the same effect foes not appear among the female respondents.

When the estimates from females for lifetime partners (fig. 1 & 2) are considered, the low estimates for the NSFG are at this point unaccounted for. In this case, data collection method does not appear to be the only driving force towards divergence as the estimates from the NSFG are far lower then the estimates from other studies which used the same data collection method. Further, there does not appear to be a difference between the reporting by women from the NCHRBS, NSW or NHSLS despite the differences in data collection method.

There are clearly many contributing factors to the differences in reporting across studies and one which has received a great deal of attention in the survey design literature in the method of administration. My findings that studies that use FTFI consistently generated higher estimates of number of partners in the last year and over the lifetime when compared to studies that used SAQs or telephone interviews are in part contradictory to much of the previous literature which suggests that SAQs elicit higher estimates of sensitive behavior. However, the GSS and NCHRBS, the only two surveys that used SAQs, are unique in other ways which may have influenced the estimates drawn from them.

The GSS variable for lifetime partners was, in fact, only partners since are eighteen. Thus, it is reasonable to assume that the actual number for lifetime is somewhat higher. Perhaps if the GSS had asked for lifetime partners, there would be little difference in responses between the GSS, which was an SAQ, and studies that used FTFI. A comparison between the NHSLS and GSS, both of which provide data on the number of partners since age eighteen, allowed me to compare results from a survey which used an SAQ and a survey that used FTFI while controlling for the question. The non-significant difference between them supports the conclusion that the GSS is an outlier in the previous analyses because of the variable used and not the mode of administration.

However, when "partners in the last year" is looked at as a top-coded variable the GSS, the only study using an SAQ to ask about partner in the last year, is still significantly different than the studies which used FTFIs.

In an attempt to parse out some of the differences between the effect of study and data collection method I would like to have been able to control for collection method in the same analysis that included study. However, because telephone surveys were only used in one survey and SAQ where only used in two surveys, each of which has its own previously mentioned idiosyncrasies, there is not enough studies to do a meaningful analysis. However, I am able to control for collection method and compare the estimates across surveys which used FTFI.

When I compare the estimates across surveys which used FTFI the results indicate that survey is not significant when both sexes are included or when just the males are

included. However, when the females are compared there is a significant difference. The likely source of this difference is the NSFG, which only includes women. This hypothesis is born out when just females from the NHSLS and NSW are compared and no significant difference is found.

When year is analyzed for trend using logistic regression the BRFS shows a significant upward trend in reporting of partners in the last year while the GSS does not. This may again be due to the construction of the variable which, in the case of the GSS, uses the medians generated from a categorical variable.

The results of these analyses suggest that study and data collection method do matter for the reporting of sexual relationships even when demographic compositions of the samples are taken into account, but perhaps not as much as intuition might suggest. The extent and direction of the effects remain unclear. The effects of survey were not as large as the effects of the demographic variables which may suggest that the impact of survey on reporting may be reduced if additional demographic characteristics such as education are included or if the characteristics used in this analysis are re-specified in order to make the groups more homogeneous. Some differences may also be due to the fact that the surveys were not all random samples despite being population based.

Additionally, my analysis either suggests that data on sexual partners is somewhat different then other types of sensitive data or presents evidence which contradicts, at least in part, some of the previous research on the effects of data collection methods on the reporting of sensitive information. Tourangeau and Smith (1996) provide an extensive list of studies which found that SAQ elicited higher estimates of sensitive behaviors than

alternative data collection methods. There own research showed ACASI and CASI to generally elicited higher levels of reporting than CASI. However, the exception in their findings is that men tended to report fewer or no more sex partners under CASI and ACASI than under CAPI. These particular findings are the most relevant because the behavior of interest is the number of sex partners. However, in my broad analysis both men and women reported significantly more partners in FTFIs than they did on SAQs. When more refined analysis were performed some of the difference between methods dropped out, but SAQ was never shown to elicit higher estimates than FTFIs.

My findings are also consistent with previous research on the differences in sexual partner reporting between men and woman. As Scott (1992) reported, men consistently reported higher numbers of partners then women. Whatever the source of this discrepancy, it does not appear to be the result of data collection method. That is to say, in the presence of another men do not appear to be prone to over-reporting while women under-report. The differences between reporting based on method of survey administration were the same for both men and women although the magnitude of the difference and the number of total partners reported were greater for males.

Conclusion

The evidence presented here does show that there is a statistically significant difference in estimates of sexual behavior across studies. However, when key differences in the studies, like sample compositions and collection method, are taken into account much of the difference disappears. Perhaps if a few other differences in the sample composition could have also been accounted for the differences would have been non-significant. The important point is that the apparent differences in the estimates seem to be a function of the survey and not the responses. The responses seem to be remarkably consistent despite the myriad differences between the studies. Further, when attempts are made to account for the differences in the studies the responses become even more consistent. This indicates that whatever bias or tendencies in reporting are effecting respondent's reporting, they are likely more uniform than previously thought, which means they can be adjusted for in future research.

Despite the fact that there were statistically significant differences across surveys it seems that respondents may be far more willing to be forthright when reporting on their sexual behavior than previously thought, and that their willingness is not dependant on the time or method of data collection. This finding bodes well for data collection in the future, as the method of administration does not appear to be impacting respondents as much as previously reported.

The results of this analysis are also promising for the future of epidemiological research and the study of disease transmission dynamics. The comparability of these data

sets suggests that different surveys may be able to serve effectively as proxies for missing classes of data in others. If this is the case it will be possible to construct second order networks from egocentric data even if egos partners are not similar to any other member of the study of which ego is a part because other surveys can be used to impute the characteristics of the alters.

Thus, while there are several differences across these surveys and the findings were statistically significant, it is my conclusion that the surveys are consistent, and that the consistency will allow researchers access to much broader pools of data.

Bibliography

Anderson, John E. and Ron Stall. "Increased reporting of male-to-male sexual activity in a national survey". *Sexually Transmitted Disease*. November 2002 643-646

Anderson RM, May RM. Infectious diseases of humans; dynamics and control, 1st ed. Oxford: Oxford University Press, 1992:234

Babbie, Sheila "The Practice of Social Research 7th ed." Wadsworth Publishing Co. 1995

Bradburn, N.M. 1983 "Response Effects." In Handbook of Survey Research, ed. P. Rossi, J. Wright, and A. Anderson, pp.289-328. New York: Academic Press

Brody, S. "Patients misrepresenting their risk factors for AIDS." *Int. J. STD. AIDS* 1995 6:392-98

Catania, J., D. R. Gibson, D. D. Chitwood, T. J. Coates. "Methodological problems in AIDS behavior research: Influences on measurement error and participation bias in studies of sexual behavior." *Psychol. Bull* 1995 108:339-62

Cantania, J., D. Binson, A. van der Straten, V. Stone. "Methodological research on sexual behavior in the AIDS era." *Annual Review of Sex Research* 1995, 6, 77-125

Carey, M. P., Carey, K. B., Maista, S. A., Gordon, C. M., and Weinhardt, L. S. "Assessing sexual risk behavior with the Timeline Followback (TLFB) approach: Continued development and psychometric evaluation with psychiatric outpatients." *International Journal of STD and AIDS* 2001, 12, 365-375

Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System User's Guide*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 1998. http://www.cdc.gov/brfss/pdf/userguide.pdf

Durant LE, Carey MP "Reliability of retrospective self-reports of sexual and nonsexual health behaviors among women." *J Sex Marital Ther*. 2002 Jul-Sep;28(4):331-8.

Hethcote, Herbert W. and James A. Yorke. Gonorrhea Transmission Dynamics and Control (Lecture Notes in Biomathmatics, vol 56) Springer Verlag, December 1984

Huttenlocher J. LV Hedges, NM Bradburn. "Reports of elapsed time: bounding and rounding processes in estimation." J. Exp. Psychol. Learn. Mem. Cognit. 1990 16:196-213

Jolly A.M., Wylie MJ. "Gonorrhea and Chlamydia care groups and sexual networks in Manitoba". *Sex Transm Infect* 2002;78(Suppl 1);i145-i151

Jones, E. and J. Forrest. 1992 "Understanding of abortion in surveys of U.S. women: 1976 to 1988." *Demography* 29:113-26.

Kalichmann S. C., Kelly, J. A., and Stevenson, L. Y. "Priming effect of HIV risk assessments on related perceptions and behavior: An experimental field study." *AIDS Behavior,* 1997, 1, 3-8.

Klovdahl, Alden S. "Networks and pathogens". *Sexually Transmitted Diseases* January 2001 Editorial.

Laumann, Edward O., John H. Gagnon, Robert T. Michael, and Stuart Michaels. The Social Organization of Sexuality "Sexual Practices in the United States." The University of Chicago Press, Ltd., London 1994

Laumann, Edward O., John H. Gagnon, Robert T. Michael, and Stuart Michaels. National Health and Social Life Survey, 1992: [United States] [Computer file]. ICPSR version. Chicago, IL: University of Chicago and National Opinion Research Center [producer], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1995.

Lewontin, R. C. Sex, lies, and social science. N.Y. Rev Books 1995 42: 24-29

McLaws, M., Oldenburg, B., Ross, M. W., and Cooper, D. A. "Sexual behavior in AIDS-related research: Reliability and validity of recall and diary measures." *Journal of Sex Research* 1990 27, 265-81

MMWR, Volume 46, Numberss-6, November 14 1997

Morris, M. "Telling tails explains the discrepancy in sexual partner reports" *Nature* Vol 365 Sept. 1993

The National Opinion Research Center. *General Social Survey 1972 - 2000 Cumulative Codebook.* 2003 http://www.icpsr.umich.edu:8080/GSS/homepage.htm

Saltzman SP, Stoddard AM, McCusker J, Moon MW, Mayer KH "Reliability of self-reported sexual behavior risk factors for HIV infection in homosexual men." *Public Health Rep*. 1987 Nov-Dec; 102(6):692-7.

Santelli, John S, Laura Duberstein Lindberg, Joyce Abma, Clea Sucoff McNeely and Michael Resnick "Adolescent Sexual Behavior: Estimates and Trends from four Nationally Representative Surveys" *Family Planning Perspectives* Vol 32(4) 2000

Schaeffer, N.C. and Presser, S. "The science of asking questions" *Annu. Rev. Sociol.* 2003. 29:65-88

Schwarz N., F. Strack, H. P. MAI. "Assimilation and contrast effects in part-whole question sequences: a conversational logical analysis." *Public Opin. Q*. 1991 55:3-23

Smith T. "Thoughts on the nature of context effects." In Context Effects in social and Psychological Research ed. N. Schwartz and S. Sudman. New York: Springer-Verlag 1992.

Smith, T., "Discrepancies between men and women in reporting number of sexual partners: A summary from four countries." *Social Biology* 39:203-11, 1992.

Sneed CD, Chin D, Rotheram-Borus MJ, Milburn NG, Murphy DA, Corby N, Fahey J. "Test-retest reliability for self-reports of sexual behavior among Thai and Korean respondents." *AIDS Educ Prev*. 2001 Aug; 13(4):302-10.

Sohler N, Colson PW, Meyer-Bahlburg HF, Susser E. "Reliability of self-reports about sexual risk behavior for HIV among homeless men with severe mental illness." *Psychiatr Serv.* 2000 Jun; 51(6):814-6.

Tanfer, K. "National Survey of Men: Design and Execution."
Family Planning Perspectives, (1993, March/April). 25(2), 83-86.

Tanfer, Koray "A User's Guide to the Machine-Readable Files and Documentation March 1995" 1991 NATIONAL SURVEY OF WOMEN Sociometrics Corporation

Taylor, J. F., Rosen, R. C., and Lieblum, S. L. "Self-report assessment of female sexual function: Psychometric evaluation of the brief index of sexual functioning for women." *Archives of Sexual Behavior*, 1994, 23, 627-43

Thomas JC, Tucker MJ. "The development and use of the concept of a sexually transmitted disease core." *J Infect Dis*; 174(suppl2):s134-43 1996

Tourangeau R, K. Rasinski, N Bradburn. "Measuring happiness in surveys: a test of the subtraction hypothesis." Public Opin. Q. 1991 55:255-66

Tourangeau, R., K. Rasinski. "Cognitive processes Underlying Context effects in attitude measurement." *Psychological Bulletin* 103:299-314 1988

Tourangeau Roger and Tom W. Smith. "Asking sensitive questions: The impact of data collection mode, question format, and question context." *Public Opinion Quarterly,* Vol. 60, Issue 2 (summer, 1996), 275-304

U.S. Dept. of Health and Human Services, National Center for Health Statistics. National Health Interview Survey, 1992: Youth Risk Behavior Supplement [Computer file]. Hyattsville, MD: U.S. Dept. of Health and Human Services, National Center for Health Statistics [producer], 1993. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1994.

U.S. Dept. of Health and Human Services, National Center for Health Statistics. National Survey of Family Growth, Cycle V, 1995 [Computer file]. ICPSR version. Hyattsville, MD: U.S. Dept. of Health and Human Services, National Center for Health Statistics [producer], 1997. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2000.

Van Duynhoven YT, Nagelkerke NJ, Van De Laar MJ. "Reliability of self-reported sexual histories: test-retest and interpartner comparison in a sexually transmitted diseases clinic." *Sex Transm Dis*. 1999 Jan; 26(1):33-42.

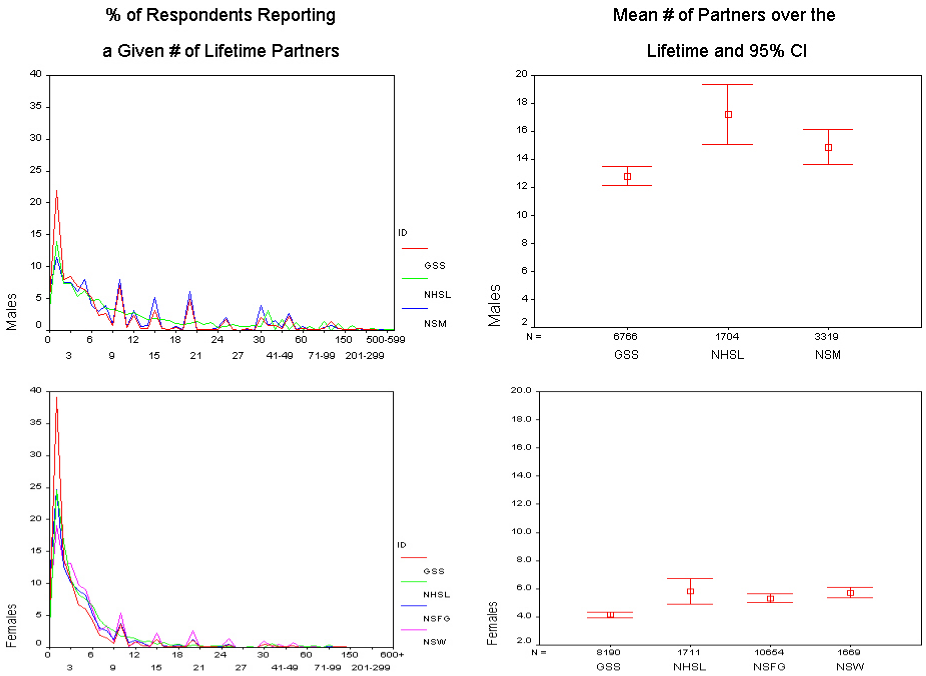
Weinhardt, Lance S., Andrew D. Forsyth, Michael P. Carey, Beth C. Jaworski and Lauren E. Durant "Reliability and Validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice" Archives *of Sexual Behavior*; 1998 Apr 27(2):155-80

Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., and Wickramasinghe, S,M. "Reliability of the timeline followback sexual behavior interview." *Annals of Behavioral Medicine*, 1998, 20, 25-30.

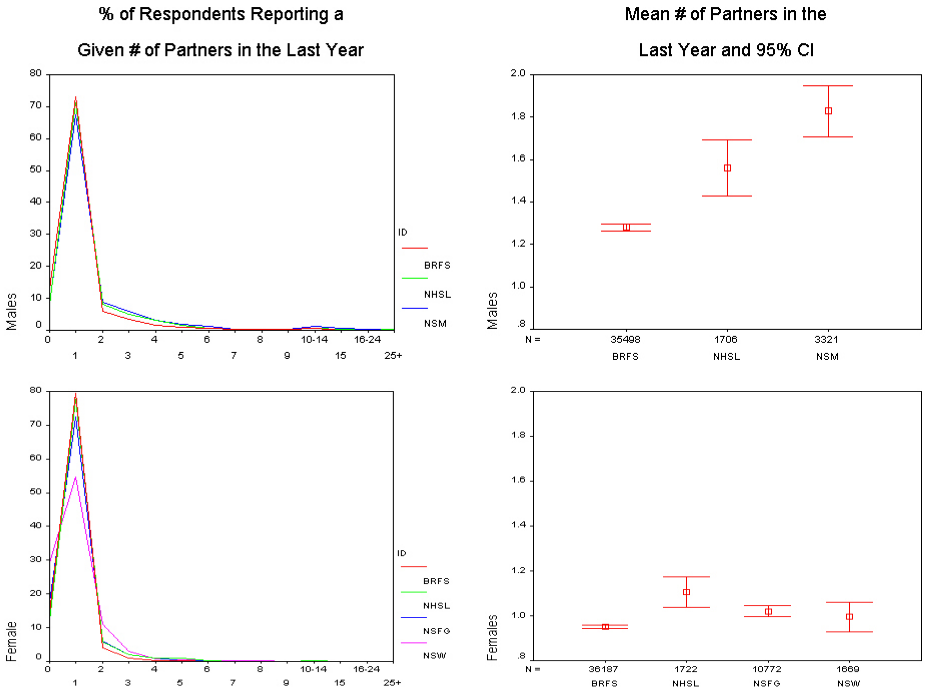
1983 National Survey of Unmarried Women, Age 20-29
A User's Guide to the Machine-Readable Files and Documentation
July 1990 Data Archive on Adolescent Pregnancy and Pregnancy Prevention, Sociometrics Corporation

1998 The National Longitudinal Study of Adolescent Health
Carolina Population Center at the University of North Carolina at Chapel Hill
Last modified 25 July 2003 **http://www.cpc.unc.edu/projects/addhealth/design.html**

Appendix A: Figures

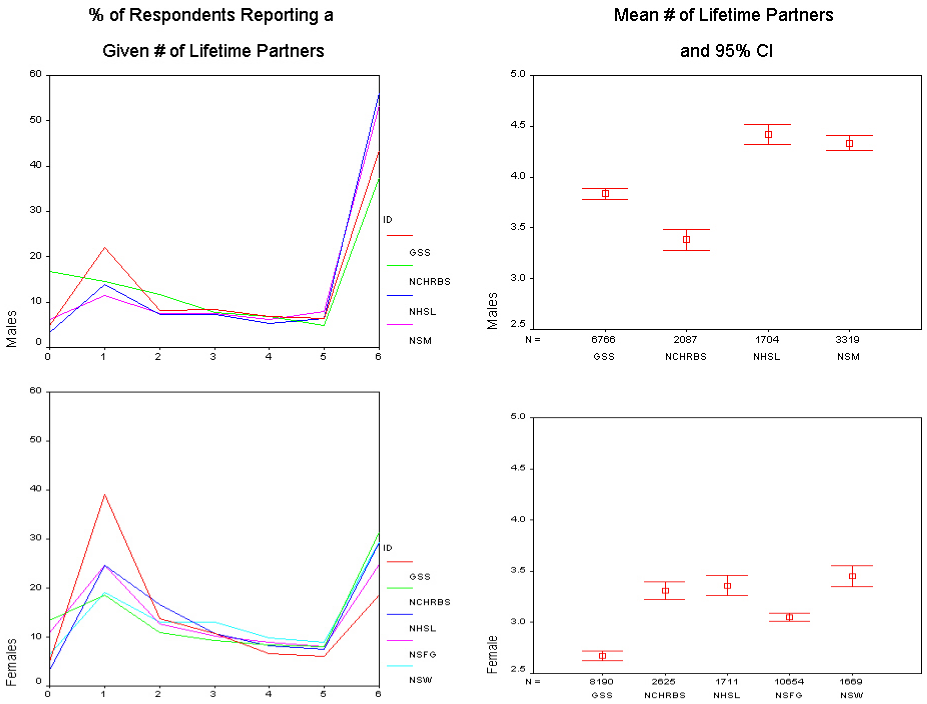


Partners in the lifetime

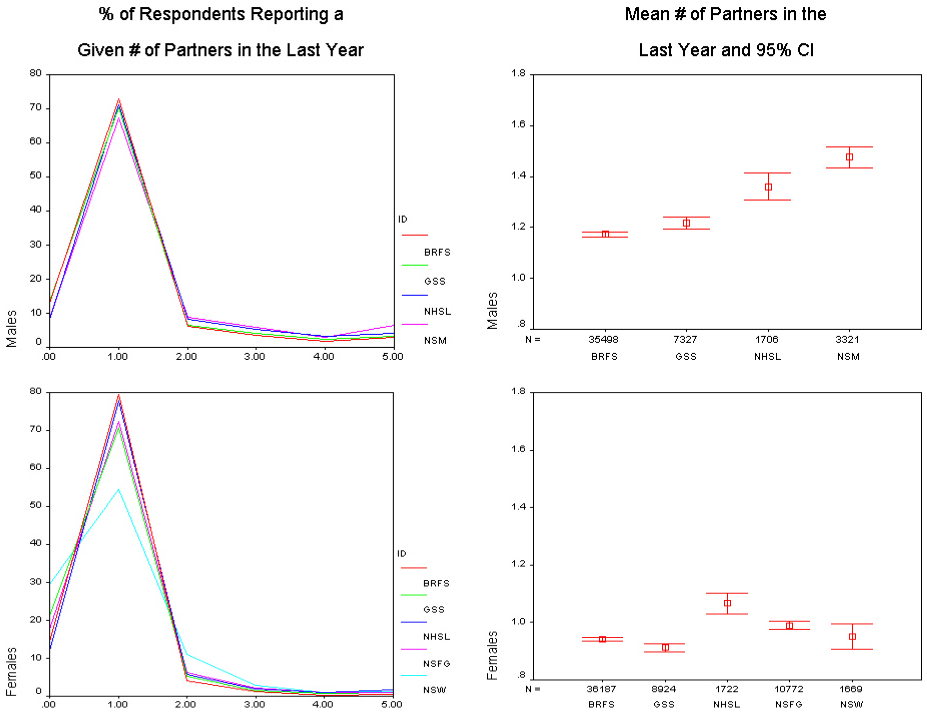


Partners in the last year

Figure 1. Line graphs and error bars for estimates of sex partners over the lifetime and in the last year reported as a continuous variable



Partners in the lifetime



Partners in the last year

Figure 2. Line graphs and error bars for estimates of sex partners over the lifetime and in the last year reported as a top-coded variable

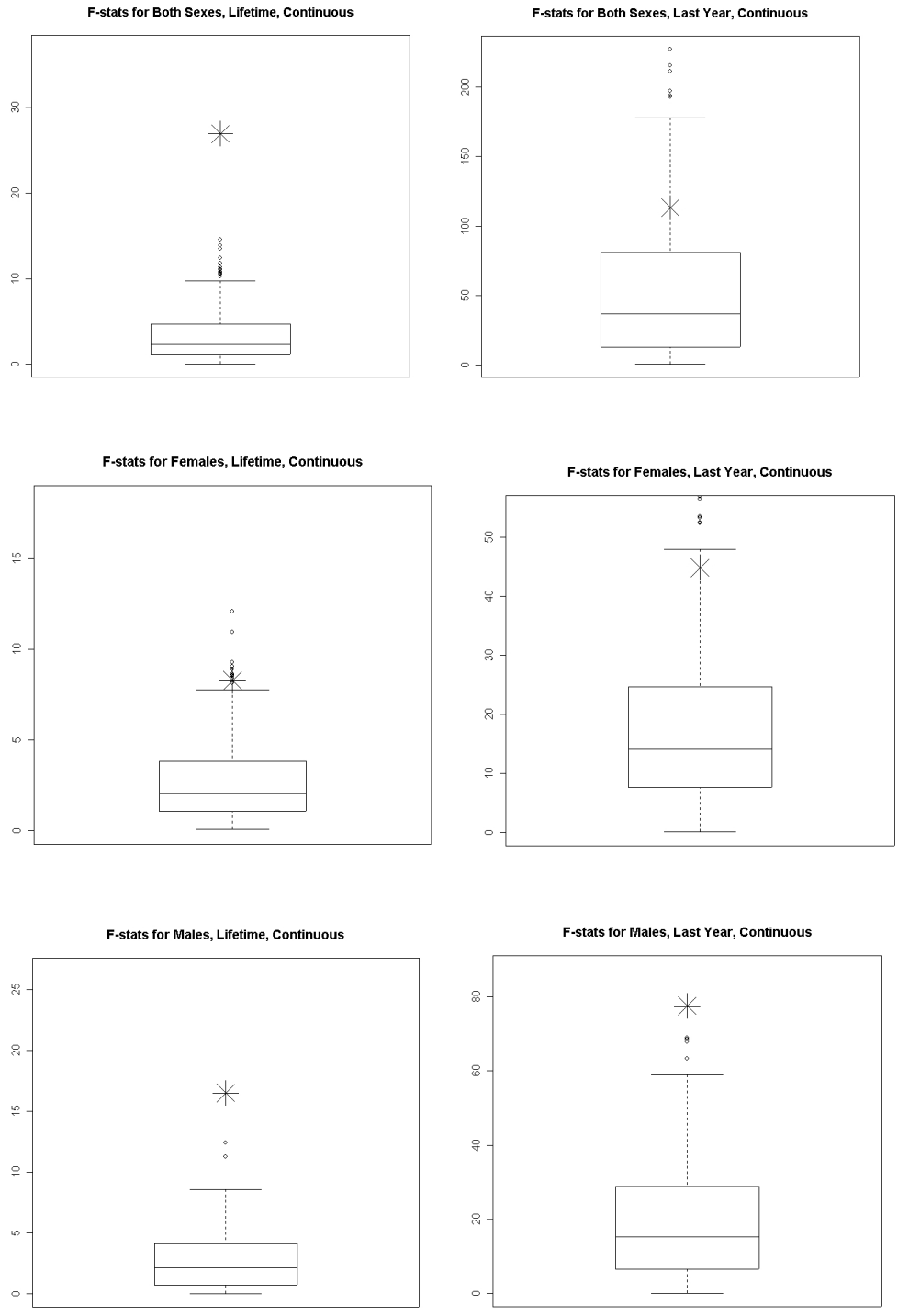


Figure 3. **Box-plots of the f-statistics generated from the permutation tests of partners in the lifetime and the last year as a continuous variable (observed marked with an asterisk).**

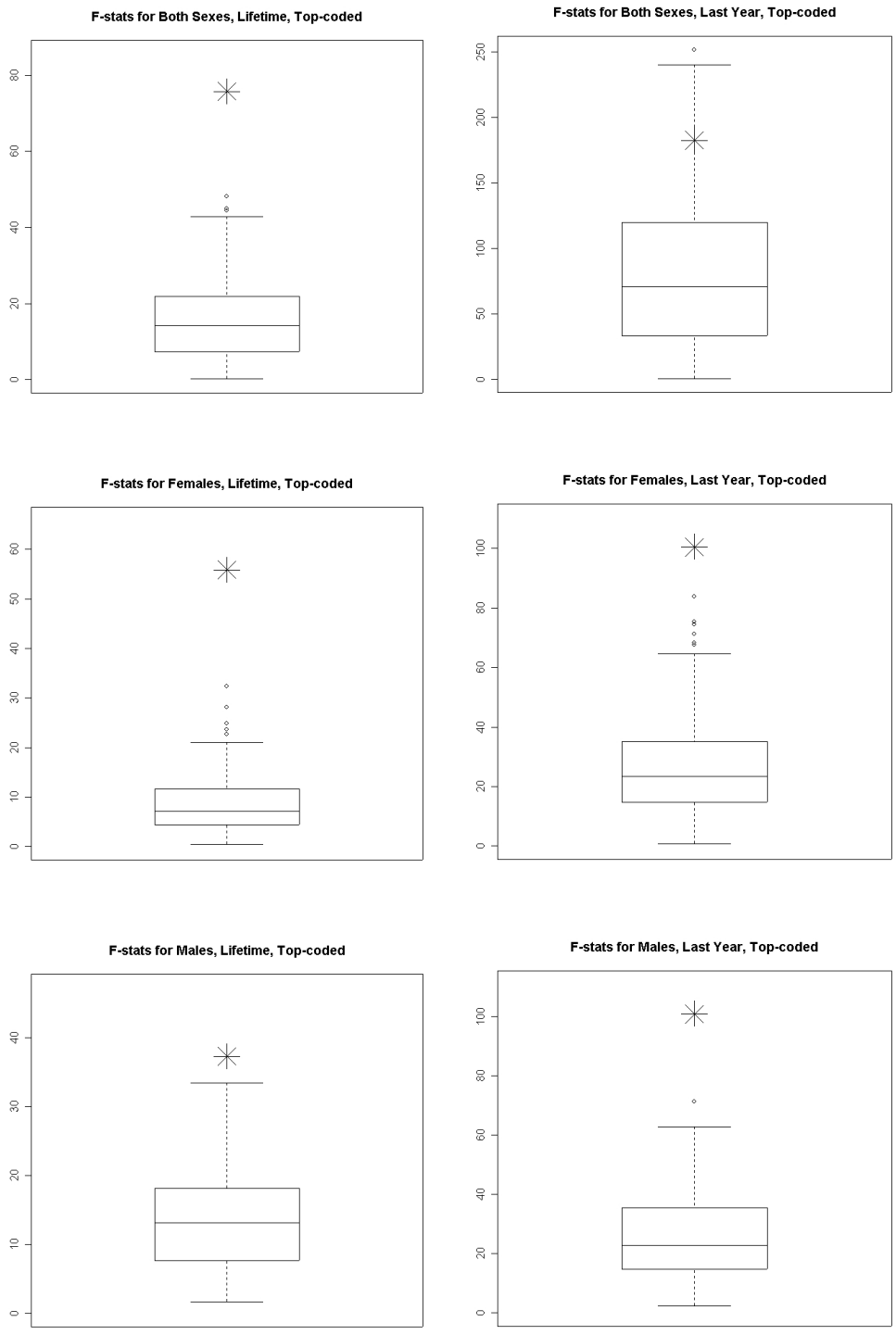


Figure 4. **Box-plots of the f-statistics generated from the permutation tests of partners in the lifetime and the last year as a top-coded variable (observed marked with an asterisk).**

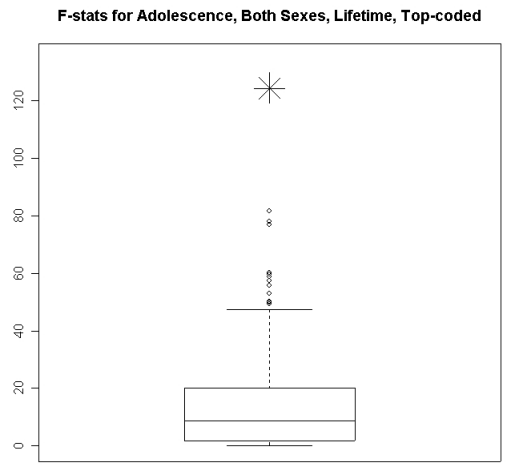


Figure 5. **Box-plot of the f-statistics generated from the permutation test of partners in the lifetime as a top-coded variable for the adolescent subgroup (observed marked with an asterisk).**

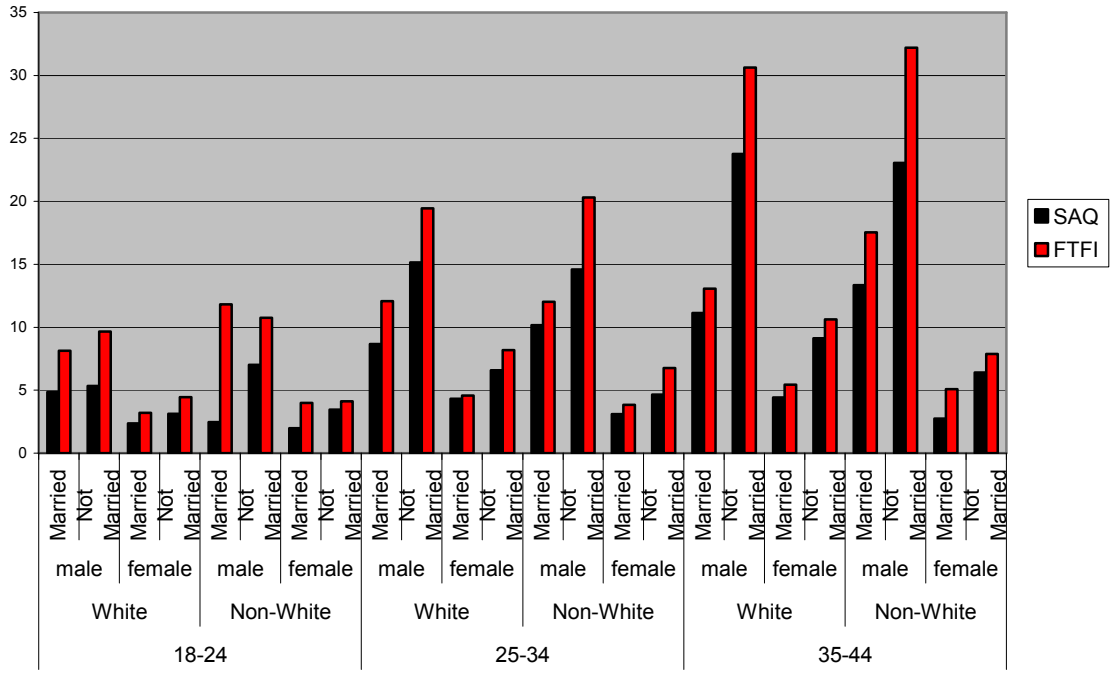


Figure 6. **Estimated number of lifetime partners as a continuous variable from surveys utilizing SAQ and surveys utilizing FTFI plotted by demographic group**

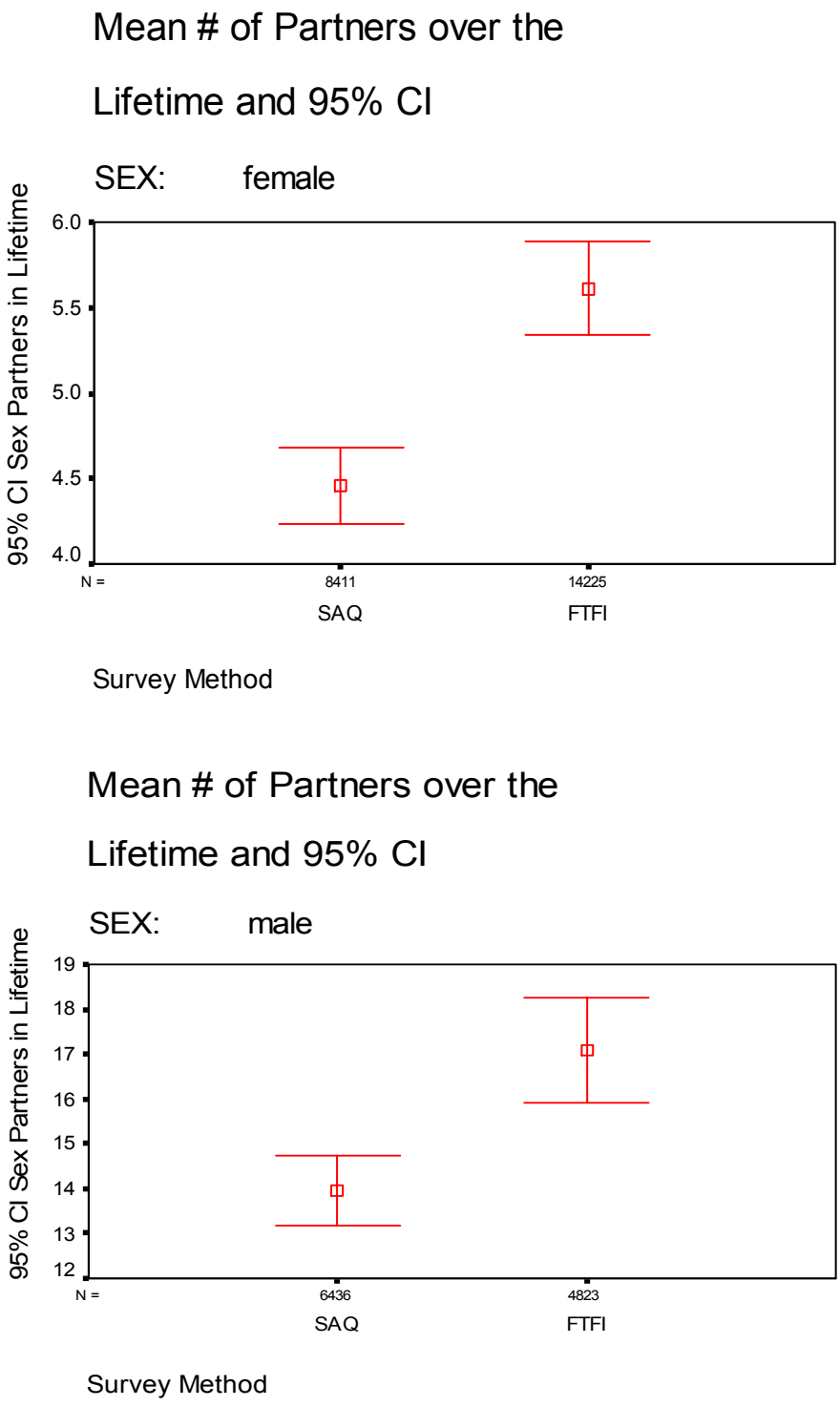


Figure 7. **The mean estimates of lifetime partners reported in surveys using SAQs and surveys using FTFI for females and males**

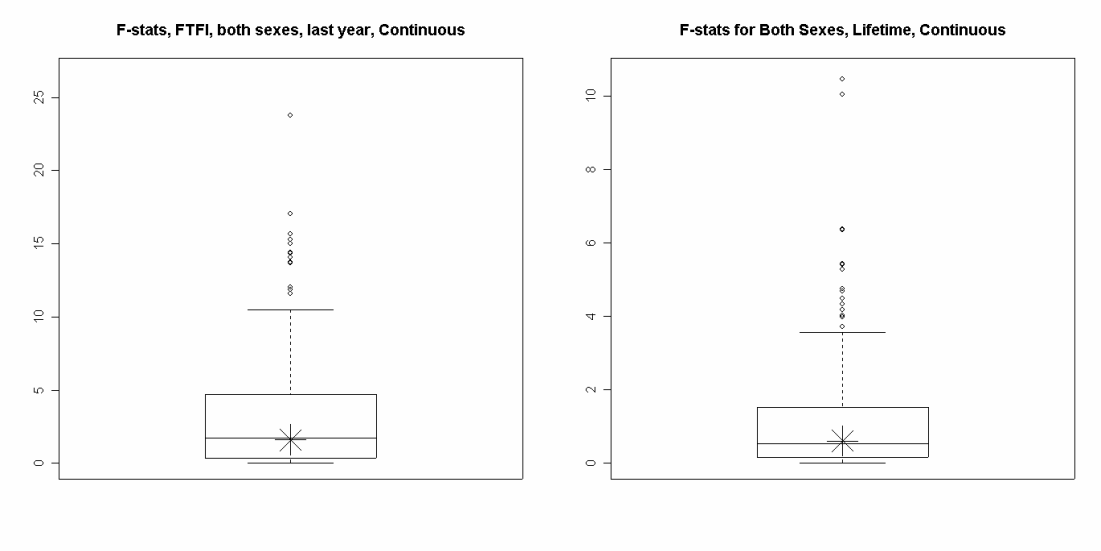


Figure 8. **Box-plots of the f-statistics generated from permutation tests of surveys which used FTFI to get estimates of partners in the lifetime and partners in the last year (observed marked with an asterisk).**

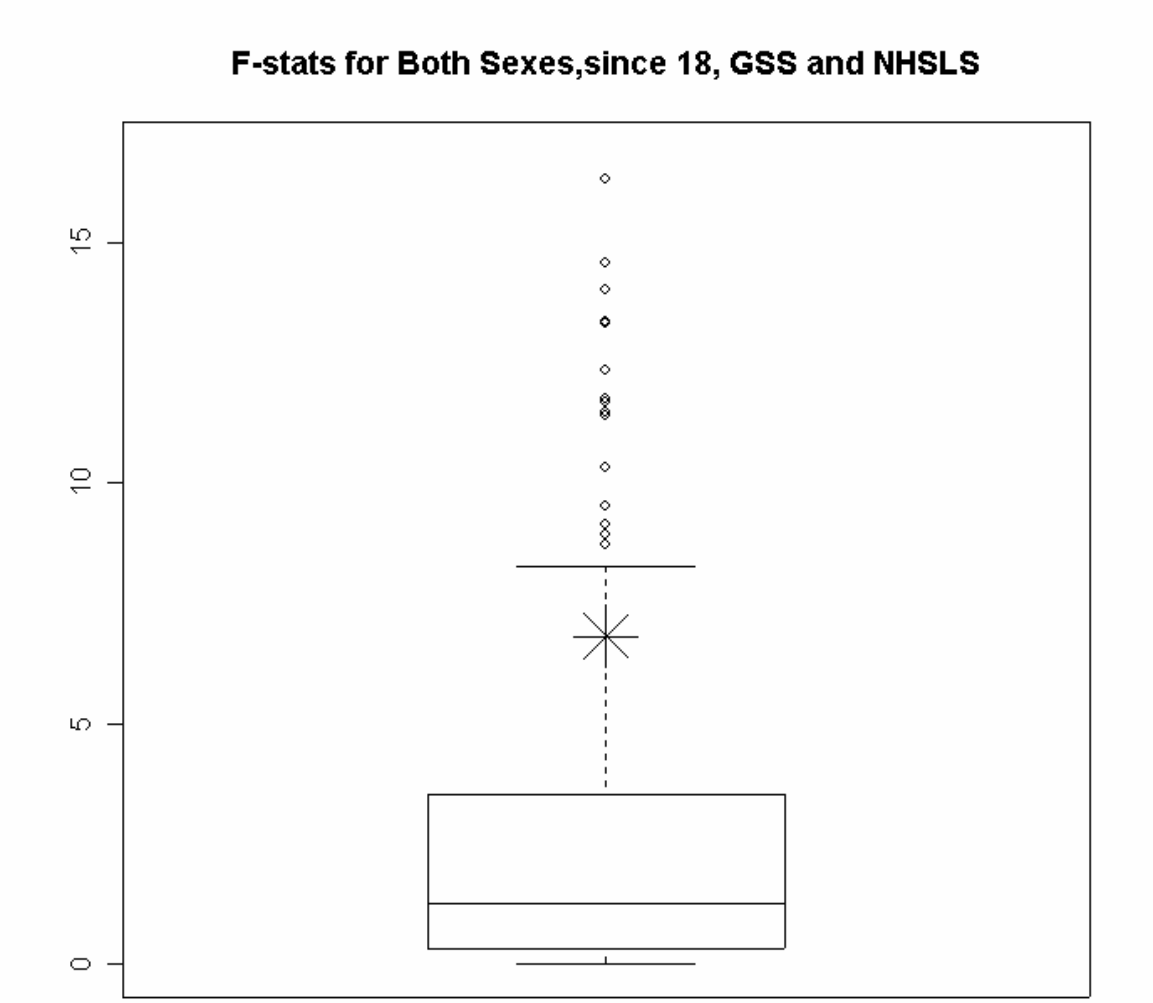


Figure 9. **Box-plots of the f-statistics generated from permutation tests of a survey which used FTFI and a survey witch used SAQ to get estimates of partners since age eighteen (observed marked with an asterisk).**

Appendix B: Tables

Table 1. An overview of the relevant attributes of the nine population-based surveys used in the study

Survey	Years	Age	Sex	Data Collection Method	Partners in Last Year	Partners in Lifetime
BRFS	1996-2000	18+	M/F	Telephone	Yes top-coded	NA
GSS	1988-91, 93, 94, 96, 98, 2000	18+	M/F	Self-administered Questionnaire	Yes Top-coded	Yes continuous
NCHRBS	1995	18+	M/F	Self-administered Questionnaire	NA	Yes Top-coded
NHSLS	1992	18-59	M/F	Face to Face Interview	Yes continuous	Yes continuous
NSFG	1995	15-44	F	Face to Face Interview	Yes continuous	Yes continuous
NSM	1991	30-39	M	Face to Face Interview	Yes continuous	Yes continuous
NSW	1991	20-39	F	Face to Face Interview	Yes continuous	Yes continuous
YRBS	1992	12-21	M/F	NA	NA	Yes Top-coded
ADD HEALTH	1994-1996	11-23	M/F	NA	NA	Yes continuous

Table 2. The questions used in each of the nine surveys and a brief description of the adjustments made to the data

Variable	Question for last year	Adjustments for last year	Question for lifetime	Adjustments for Lifetime
BRFS	During the past twelve months, with how many people have you had sexual intercourse?	No adjustments. Responses were top-coded at 76+. 3 respondents reported 76+ partners.	NA	NA
GSS	How many sex partners have you had in the last 12 months?	Responses were categorical for values greater then 4 so all responses greater then 4 were coded as 5. These data were only used in the truncated analysis.	Now thinking about the time since your 18th (including the past 12 months) how many male partners have you had sex with? Now thinking about the time since your 18th (including the past 12 months) how many female partners have you had sex with?	Constructed by adding number of male partners since 18 and number of female partners since 18.
NCHRBS	NA	NA	During your life, with how many females have you had sexual intercourse? During your life, with how many males have you had sexual intercourse?	Constructed by adding the number of male partners in lifetime and number of female partners in lifetime.
NHSLS	Thinking back over the past 12 months, how many people, including men and women, have you had sexual activity with, even if only one time?	NA	See Laumann et al 1994 appendix C	For a description contact the author

Table 2. (continued)

NSFG	During the last 12 months, that is, since (MONTH /YEAR), how many men, if any, have you had sexual intercourse with? Please count every male sexual partner, even those you had sex with only once. (Probe for range if R is unable to recall exact number.)	Constructed from hi-lo estimates	Counting all your male sexual partners, even those you had intercourse with only once, how many men have you had sexual intercourse with in your life? (Probe for range if R is unable to recall exact number.)	Raw number constructed from hi low estimates
NSM	Since January 1990, how many different women have you had vaginal intercourse with? Since January 1990, how many different partners have you had anal sex with?	Constructed from the greater of vaginal sex partners in 1990 and anal sex partners in 1990	With how many different women have you ever had vaginal intercourse? With how many partners have you ever had anal intercourse?	Constructed from the greater of vaginal sex partners in lifetime and anal sex partners in lifetime
NSW	With how many different men did you have vaginal intercourse since January 1990? With how many different men did you have vaginal intercourse since January 1990?	Constructed as the greater of vaginal sex partners in 1990 or anal sex partners in 1990	With how many different men have you ever had vaginal intercourse? With how many different men have you ever had anal intercourse?	Constructed from the greater of vaginal sex partners in lifetime or anal sex partners in lifetime