Sexually Transmitted Diseases, October 2004, Vol. 31, No. 10, p.596–600
DOI: 10.1097/01.olq.0000140011.93104.32

Studies Relying on Passive Retrospective Cohorts Developed From Health Services Data Provide Biased Estimates of Incidence of Sexually Transmitted Infections

CHARLOTTE K. KENT, MPH,* JANICE K. CHAW, MPH,* ROBERT P. KOHN, MPH,* YING Q. CHEN, PHD,† AND JEFFREY D. KLAUSNER MD, MPH*

Objective: **Passive retrospective cohorts composed of persons who have tested 2 or more times for a sexually transmitted infection (STI) of interest during clinical visits have been used to estimate STI incidence. We hypothesized that the analytic period of a passive cohort might affect the estimate of STI incidence, with shorter periods yielding higher estimates of incidences of infection.**

Study: **We analyzed data collected from women, 12 to 24 years of age, tested for chlamydia 2 or more times at 6 sites in San Francisco between January 1997 and December 2000. Incidence was calculated for 10 different analytic periods.**

Results: **The calculated incidence of chlamydial infection during 1997 was 16.8 (95% confidence interval [CI], 10.9–24.0) per 1000 person-months of follow up. The calculated incidence dropped markedly as the analytic period lengthened, with the incidence estimated to be 9.7 (95% CI, 8.6–10.9) using a study period of 4 years (1997–2000). Estimates of incidence were similar when using the same analytic period, regardless of calendar year, and there was a similar decline in estimated incidence using longer analytic periods.**

Conclusions: **Estimates of STI incidence based on passive cohort data may have limited epidemiologic value because incidence measures may be highly dependent on the analytic period.**

*From the *Sexually Transmitted Disease Prevention and Control Services, San Francisco Department of Public Health, San Francisco, California; and the †School of Public Health, Division of Biostatistics, University of California Berkeley, Berkeley, California*

SEXUALLY TRANSMITTED INFECTIONS (STIs) such as *Chlamydia trachomatis*, *Neisseria gonorrhoeae,* and HIV cause numerous adverse sequelae. Chlamydial and gonococcal infections increase risk for HIV transmission and can lead to infertility and ectopic pregnancy.[1,2] STIs are common and costly. Chlamydia and gonorrhea are the 2 most commonly reported infections in the United States, with 783,242 chlamydia infections and 361,705 gonococcal infections reported in 2001.[3] Chlamydial infections resulted in an estimated $1.9 billion in direct medical costs in 1998.[4] Direct medical costs for HIV infection in the United States were estimated to be $6.7 billion in 1996.[5]

Because of the sequelae and high cost of STIs, accurate estimates of incidence are important to monitor trends, develop and evaluate prevention strategies, make policy recommendations, and allocate resources. However, it its difficult to interpret incidence trends based only on reported cases because STIs such as chlamydial, gonococcal, and early HIV infections are primarily asymptomatic.[1,2] Therefore, detection of most infections relies on screening of asymptomatic persons. The extent to which reported chlamydia, gonorrhea, and HIV cases represent the true level of infection in the community is dependent on how comprehensive screening is in the population at risk for infection.[6]

Developing estimates of incidence from prospective cohort studies would provide the most accurate estimates of the incidence of STIs, but this type of study is very costly to conduct and impractical. As an alternative to prospective cohort studies, several investigators have developed passive retrospective cohorts to estimate incidence.[7–20] Generally, data to analyze incidence in these studies were collected for other purposes, making these studies inexpensive to conduct. These passive cohorts are composed of persons who have been tested 2 or more times for the STI of interest in defined settings. For example, Burstein and colleagues,[8] using clinic-based data from Baltimore, found a high incidence of chlamydial infection in women younger than 25 years of age. Based on their findings, they recommended that sexually active young women be screened for chlamydia twice a year[8] rather than once a year as currently recommended.[21]

However, studies of passive retrospective cohorts based on clinic data may provide biased estimates of incidence, because the reasons why a person returns for a clinic visit may not be independent from his or her risk of an STI. Individuals may have symptoms, know they were a sexual contact of a person with an STI, or have had a sexual encounter that puts them at risk for an infection and thus seek STI testing. For example, a prospective cohort study found, not surprisingly, that men's HIV testing behavior was significantly related to their risk for acquiring HIV infection, with persons at greatest risk testing most often.[22] In addition, studies from San Francisco[11] and London[17] found that the highest users of repeat HIV testing were persons with the highest risk behaviors and the highest incidence of HIV. In contrast, in active prospective cohort studies with high follow-up rates and nondifferential loss to follow up, the risk for an STI and the reason for a return visit theoretically are independent, and unbiased estimates of risk and incidence can be obtained.

Given the potential selection bias that could occur in passive

This research was supported by the San Francisco Department of Public Health.

Correspondence: Charlotte Kent, STD Prevention and Control Services, 1360 Mission St., Suite 401, San Francisco, CA 94103. E-mail: Charlotte.kent@sfdph.org.

Received for publication December 18, 2004, and accepted May 6, 2004.

retrospective cohort studies of STIs described here, we hypothesized that the study period of a passive cohort might affect the estimate of STI incidence, with shorter study periods yielding higher estimates of incidences of infection. We analyzed data collected from women tested for chlamydia in San Francisco to test this hypothesis.

Methods

Study Population

As a component of routine surveillance, the San Francisco Department of Public Health, STD Prevention and Control Services conducts prevalence monitoring for chlamydia using data collected on all males and females tested for chlamydia at 6 sites in San Francisco: 2 family planning clinics, the municipal sexually transmitted disease (STD) clinic, 1 teen clinic, and the youth and adult detention facilities. For this study, we examined data collected during chlamydia testing at these sites of women, 12 to 24 years of age, between January 1997 and December 2000. Subjects were those women with more than 1 visit, more than 30 days apart, at any 1 of the 6 sites. The return visit was not necessarily to the same site as the initial visit, but had to be to 1 of the 6 sites to be captured in our data system. Demographic and clinical data were collected on a standardized form. Urine specimens were tested for chlamydia using a nucleic acid amplification test (NAAT) (LCx; Abbott Laboratories) or ProbeTec (BD Laboratories).

Data Analyses

Crude Incidence Measures. We defined an incident infection of chlamydia as the first positive chlamydia NAAT test at least 30 days after a previous negative or positive NAAT that presumably had been treated during the study period, as described by Burstein and colleagues.[7,8] Because we could not verify treatment of all women with positive chlamydia NAAT tests, we also examined incident infections only among women with a previous negative test. We calculated incidence as the rate of incident infections per person-month of follow up. We calculated 95% confidence intervals for the incidence rate under the assumption of an exponential distribution for time to disease using a maximum likelihood estimate of variance.[23] Person-months of follow up for women with incident infections were calculated as the sum of the time between their baseline visit and the midpoint between their last negative test and their first positive test (the analytic end point). For persons who were negative in all subsequent tests, person-months were calculated as the sum of time between their first test and their last test. Incidence was calculated for 10 different study periods: 1-year study periods in calendar years 1997, 1998, 1999, and 2000; 2-year periods in calendar years 1997–1998, 1998–1999, and 1999–2000; 3-year periods in calendar years 1997–1999 and 1998–2000; and a 4-year study period in calendar years 1997–2000.

Median Time to Infection Calculations. In a method similar to Burstein et al.,[7,8] we estimated the time to incident infection by calculating the median, 25%, and 75% quartile time intervals in months as the sum of the time between their baseline visit and the midpoint between their last negative test and their first positive test during the 10 study periods described here. (Burstein et al. estimated the time to incident infection by summing the time between tests rather than taking the midpoint.)

Poisson Regression Analyses. Because increasing age is strongly associated with a decreasing risk of chlamydial infection,[3,24] and race/ethnicity also is a strong predictor of chlamydial infection,[3,24] we examined the effect of these covariates on incidence through Poisson regression modeling using Statistical Analysis Software (SAS) version 8 (SAS Institute, Cary, NC). We modeled the role of age group (12–16 years, 17–20 years, and 21–24 years) and race/ethnicity on incidence by study periods of 1 year: 1997, 1998, 1999, and 2000. In addition, we modeled the role of these covariates on estimating incidence by varying analytic periods beginning in 1997: 1 year, 2 years, 3 years, and 4 years.

Results

Population Characteristics

Between January 1997 and December 2000, 2165 women 12 to 24 years of age had more than 1 visit with chlamydia testing at 1 of the 6 participating screening venues. Among these women, there were 6641 visits greater than 30 days apart (mean, 3.1 visits/women; range, 2–14). The overall prevalence of chlamydia among all visits was 9.6%, but varied significantly by site ($P < 0.001$). Most first visits were made at the detention facilities (41.3%) and the municipal STD clinic (33.0%), which were among the highest prevalence sites. The mean and median age of our study population was 19 years. As expected, the prevalence of chlamydial infection at first visit decreased significantly with increasing age ($P < 0.001$) (Table 1). The race/ethnicity of those tested was diverse, and the prevalence of infection at first visit varied significantly by race/ethnicity ($P < 0.001$) (Table 1). Among the 6651 visits with chlamydia testing, 95.3% of tests were performed using LCx and the remaining tests were performed using ProbeTec.

Crude Incidence Measures

We calculated the incidence of chlamydial infection in women with either a previous negative or positive test beginning in calendar year 1997 with analytic periods lasting 1 year, 2 years, 3 years, and 4 years. The calculated incidence of chlamydial infection was 16.8 (95% confidence interval [CI], 10.9–24.0) per 1000 person-months of follow up using observations obtained during one calendar year (Table 2). The calculated incidence dropped markedly as the analytic period lengthened, with the incidence estimated to be 9.7 (95% CI, 8.6–10.9) infections per 1000 person-months using an analytic period of 4 years (Table 2). To determine if this decreasing incidence was an artifact of decreasing community prevalence, we estimated the incidence by varying analytic periods beginning in 1998, 1999, and 2000 (Table 2). Figure 1 demonstrates that the estimates of incidence were similar when using the same length of analytic period, regardless of the calendar year with which we began the analysis, and that there was a similar decline in estimated incidence using a longer analytic period.

The incidence trends described in Table 2 and Figure 1 were seen when we calculated the incidence of chlamydial infection among only those women with a previous negative test. However, the measures of incidence were lower (data not shown).

Median Time to Infection

Next, we examined if there would be a lengthening of median time to incident chlamydial infection as the analytic period lengthened. Beginning in 1997, with a 1-year analytic period, the median time to incident infection was 2.7 months (25th quartile 1.7 months, 75th quartile 3.7 months) (Table 3). With an analytic period of 4 years, the median time to incident infection increased to 6.6 months (25th quartile 2.8 months, 75th quartile 11.1 months) (Table 3). The median time to infection was similar when the analytic period was the same length, regardless of which calendar

TABLE 1. Selected Characteristics at First Visit of a Passive Cohort of Women 12–24 Years of Age Tested on 2 or More Occasions for Chlamydia in San Francisco, 1997–2000 (N = 2,165)

Characteristic	Proportion with Characteristic	Percentage Positive for Chlamydia	*P*-value*
Screening site			<0.0001
STD clinic	33.0%	11.3%	
Adult detention	21.8%	13.4%	
Youth detention	19.5%	19.0%	
Teen clinic	14.6%	6.4%	
Family planning clinic 1	8.1%	23.4%	
Family planning clinic 2	3.1%	17.7%	
Age-group (years)			<0.0001
12–16	24.4%	18.9%	
17–20	45.6%	13.7%	
21–24	30.0%	9.4%	
Race/ethnicity			<0.0001
African-American	46.7%	18.0%	
Asian/Pacific Islander	14.7%	14.9%	
Hispanic	15.5%	7.5%	
White	21.5%	8.7%	
Other	1.5%	6.3%	
Test			0.126
LCx	95.3%	13.5%	
ProbeTec	4.7%	18.8%	

* Two-sided *P*-values from chi-squared tests comparing prevalence of chlamydia by selected characteristics.

year the study period began, and there was a consistent pattern of lengthening of median time to incident infection, the longer the analytic period (Table 3).

Poisson Regression Analyses

To control for the possible confounding effects of age and race/ethnicity, we modeled their effect on the incidence of chlamydial infection for an analytic period of 1 year for 4 calendar years using Poisson regression. The rate ratios comparing 1997, 1998, or 1999 with 2000 were not significantly different from 1.0 (all *P* values >0.52), consistent with data in the rows of Table 2. Race/ethnicity also was not significantly associated with the incidence of chlamydial infection after controlling for age (all *P* values >0.29). The final model included age groups only, and the incidence of infection was 1.9 times greater ($P = 0.009$) among females 12 to 16 years compared with females 21 to 24 years, and 1.2 times greater ($P = 0.47$) among females 17 to 20 years compared with females 21 to 24 years.

Using Poisson regression, we then modeled the effect of age and race/ethnicity on incidence of chlamydial infection by 4 increasing analytic periods, with all analytic periods beginning in calendar year 1997. Race/ethnicity was not significantly associated with incidence in this model either. However, length of study period and age group remained associated with incidence of chlamydial infection. Compared with a study period of 4 years, the incidence rate ratio (RR) for a 1-year study period was 39.0 (95% CI, 12.5–121.7), for 2 years RR = 7.1 (95% CI, 2.2–22.7), and for 3 years RR = 2.3 (95% CI, 0.67–7.8). In addition, compared with females 21 to 24 years, females 12 to 16 years had an incidence rate ratio of 3.5 (95% CI, 2.5–4.7) and females 17 to 20 years had a RR = 1.9 (95% CI, 1.4–2.6).

Discussion

Using passive retrospective cohort data, we observed that the incidence of chlamydial infection among women younger than 25 years decreased with increasing duration of the study period, even after adjusting for age. The declining incidence with increasing analytic period could not be attributed to declines in the incidence of infection in the population studied over time, because the incidence was stable over varying calendar years when the duration of study was held constant. Given no apparent change in population incidence, the estimates of incidence should be similar regardless of analytic period if the processes for collecting data are

TABLE 2. Incidence* of Chlamydial Infection (with 95% confidence intervals) by Year Began Analytic Period and Length of Analytic Period Among a Passive Cohort of Women 12–24 years of Age at Initial Visit: San Francisco, 1997–2000

Length of Analytic Period (years)	Beginning Year of Analytic Period			
	1997	1998	1999	2000
1	16.8 (10.9–24.0)	20.8 (14.4−28.3)	16.4 (10.6−23.5)	20.0 (12.9−28.6)
2	13.4 (11.0−16.0)	12.9 (10.6−15.4)	13.8 (11.3−16.6)	–
3	10.9 (9.4−12.4)	11.1 (9.6−12.7)	–	–
4	9.7 (8.6−10.9)	–	–	–

*Cases per 1,000 person months.

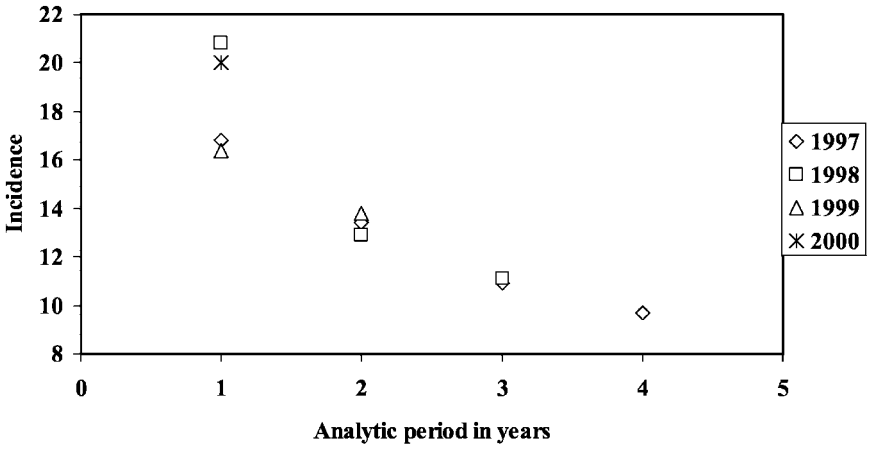

Figure 1. Incidence of chlamydail infection* among women younger than 25 years of age by calendar year and analytic period of the passive cohort: San Francisco, 1997-2000

*per 1000 person-months of follow-up

unbiased. Thus, our findings show that analyses based on passive cohort data may provide biased estimates of incidence of STIs.

A plausible explanation for these findings is that persons at greater risk for an STI return in a shorter period of time than persons at lower risk, resulting in a selection bias of those included in passive retrospective cohort data. Several recent studies examining HIV testing behaviors have found that overall, the more often persons tested for HIV, the greater was their risk behavior and the higher was their incidence of HIV infection.[11,17,19] Thus, seeking STI testing and care is not independent of risk of STI infection.

Burstein and colleagues[8] created a passive retrospective cohort of young women from data collected over a 33-month period. Based on median time to infection in this passive cohort, they recommended that young women be screened for chlamydial infection every 6 months.[8] If Burstein et al. had used data collected over a shorter or longer period, it is likely that their findings would have been different. For example, in our analyses, the median time to an incident chlamydial infection was highly dependent on the analytic period; ranging from 3 months with an analytic period of 1 year to 7 months with an analytic period of 4 years. The lengthening median time to infection with longer follow up is directly related to the declining observed prevalence in this biased data. Furthermore, our disparate findings illustrate that it may be inappropriate to use passive retrospective cohort data to develop policy recommendations about frequency of screening. Prospective cohort studies would provide the most valid data for developing such recommendations.

Several investigators have used passive retrospective cohort data to examine HIV incidence trends over time.[12,18,19] We found that when we used the same study duration over different calendar periods, we had a similar incidence of chlamydial infection, which is consistent with other data from San Francisco.[25] Although passive cohort data may not provide a valid estimate of incidence in the population of interest, this type of data may elucidate valid trends in incidence if there are no changes in why persons seek testing. However, consistent testing behaviors over time among persons in the passive cohort would be critical to use this type of data for trend analyses. An example of when testing behaviors might change over time include recent HIV prevention campaigns in San Francisco to promote HIV testing every 6 months for all sexually active gay and bisexual men, regardless of their actual risk of infection (Terry Dowling, HIV Prevention Section, San Francisco Department of Public Health, personal communication, 2003). If these HIV testing campaigns were successful in increasing the frequency of testing in lower-risk gay and bisexual men, from passive cohort data it would appear that the incidence of HIV had fallen, even if there was no change in the true overall incidence among gay and bisexual men.

Although studies based on passive retrospective cohort data are easy and inexpensive to conduct, their epidemiologic value may be limited given the substantial concerns raised here regarding the validity of their findings. The amount of bias in incidence estimates depends on the degree of association between the exposures and the likelihood of returning to the clinic. It is likely that the degree of bias varies by testing venue, ie, there may be a greater association between exposure and returning to the clinic for persons tested at an STD clinic than at a family planning clinic. The rate of decay in high-risk behaviors over time also is likely to vary depending on where persons are tested, eg, youth detention setting versus school-based clinics. Therefore, there needs to be careful

TABLE 3. Median Time in Months (25% and 75% quartiles) to Incident Chlamydial Infection by Year Began Analytic Period and Length of Analytic Period Among a Passive Cohort of Women 12–24 Years of Age at Initial Visit: San Francisco, 1997–2000

Length of Analytic Period (years)	Beginning Year of Analytic Period			
	1997	1998	1999	2000
1	2.7 (1.7–3.7)	2.8 (1.7–4.2)	1.9 (1.2–3.5)	2.1 (1.0–2.3)
2	4.2 (2.6–7.2)	4.0 (2.3–7.5)	3.9 (1.9−7.1)	–
3	5.4 (2.8–10.4)	5.6 (2.8–9.3)	–	–
4	6.6 (2.8−11.1)	–	–	–

consideration of the population being screened in interpreting results. Further research could help quantify the strength of this association for patients with STIs and therefore help estimate the bias that might be present in the estimates by population and clinic type. Because our hypothesis is not likely particular to STDs, further research on this phenomenon should be in the mathematics of the rates estimates and how they may be affected by the bias we have identified. This should eliminate concerns that might be particular to STIs or the population we used in this study.

References

1. Stamm WE. *Chlamydia trachomatis* infections of the adult. In: Wasserheit JN, ed. Sexually Transmitted Diseases, 3rd ed. San Francisco: McGraw-Hill; 1999:407–422.
2. Hook EW, III, Handsfield HH. Gonococcal infections in the adult. In: Wasserheit JN, ed. Sexually Transmitted Diseases, 3rd ed. San Francisco: McGraw-Hill; 1999:451–466.
3. Sexually Transmitted Disease Surveillance, 2001. Atlanta: Department of Health and Human Services, Centers for Disease Control and Prevention; 2002.
4. Rein D, Kassler W, Irwin K, Rablee L. Direct medical costs of pelvic inflammatory disease and its sequelae: Decreasing, but still substantial. Obstet Gynecol 2000; 95:397–402.
5. Bozzette SA, Berry SH, Duan N, et al. The care of HIV-infected adults in the United States. N Engl J Med 1998; 339:1897–1904.
6. Klausner J, McFarland W, Bolan G, et al. Knock-knock: A population-based survey of risk behavior, health care access, and *Chlamydia trachomatis* infection among low-income women in the San Francisco Bay Area. J Infect Dis 2001; 183:1087–1092.
7. Burstein G, Gaydos C, Diener-West M, Howell M, Zenilman J, Quinn T. Incident *Chlamydia trachomatis* infections among inner-city adolescent females. JAMA 1998; 280:521–526.
8. Burstein GR, Zenilman JM, Gaydos CA, et al. Predictors of repeat *Chlamydia trachomatis* infections diagnosed by DNA amplification testing among inner city females. Sex Transm Infect 2001; 77:26–32.
9. Burstein GS, MH, et al. Adolescent chlamydia testing practices and diagnosed infections in a large managed care organization. Sex Transm Dis 2001; 28:477–483.
10. Rietmeijer CA, van Bemmelen R, Judson FN, Douglas JM Jr. Incidence and repeat infection rates of *Chlamydia trachomatis* among male and female patients in an STD clinic: Implications for screening and rescreening. Sex Transm Dis 2002; 29:65–72.
11. Fernyak SE, Page-Shafer K, Kellogg TA, McFarland W, Katz MH. Risk behaviors and HIV incidence among repeat testers at publicly funded HIV testing sites in San Francisco. J Acquir Immun Defic Syndr 2002; 31:63–70.
12. Kellogg T, McFarland W, Katz M. Recent increases in HIV seroconversion among repeat anonymous testers in San Francisco. AIDS 1999; 13:2303–2304.
13. Kellogg TA, Clements-Nolle K, Dilley J, Katz MH, McFarland W. Incidence of human immunodeficiency virus among male-to-female transgendered persons in San Francisco. J Acquir Immun Defic Syndr 2001; 28:380–384.
14. Kellogg TA, McFarland W, Perlman JL, et al. HIV incidence among repeat HIV testers at a county hospital, San Francisco, California, USA. J Acquir Immun Defic Syndr 2001; 28:59–64.
15. McFarland W, Kellogg TA, Dilley J, Katz MH. Estimation of human immunodeficiency virus (HIV) seroincidence among repeat anonymous testers in San Francisco. Am J Epidemiol 1997; 146:662–664.
16. McFarland W, Kellogg TA, Louie B, Murrill C, Katz MH. Low estimates of HIV seroconversion among clients of a drug treatment clinic in San Francisco, 1995 to 1998. J Acquir Immun Defic Syndr 2000; 23:426–429.
17. Elford J, Leaity S, Lampe F, et al. Incidence of HIV infection among gay men in a London HIV testing clinic, 1997–1998. AIDS 2001; 15:650–653.
18. del Romero J, Castilla J, Garcia S, Clavo P, Ballesteros J, Rodriguez C. Time trend in incidence of HIV seroconversion among homosexual men repeatedly tested in Madrid, 1988–2000. AIDS 2001; 15:1319–1321.
19. McDonald A, Donovan B, O'Conner C, et al. Time trends in HIV incidence among homosexually active men seen at sexual health clinics in Australia, 1993–1999. J Clin Virol 2001; 22:297–303.
20. Kang M-S, Chow JM, Dunn T, Deosaransignh K, Black S, Bolan G. Re-infection with *Chlamydia trachomatis* in a large Northern California HMO: Implications for screening. 2002 National STD Prevention Conference; San Diego, CA; 2002.
21. Centers for Disease Control and Prevention. Sexually transmitted disease treatment guidelines, 2002. MMWR Morb Mortal Wkly Rep 2002; 51(RR-6):1–78.
22. MacKellar DA, Valleroy LA, Secura GM, et al. Repeat HIV testing, risk behaviors, and HIV seroconversion among young men who have sex with men. J Acquir Immun Defic Syndr 2002; 29:76–85.
23. Hirsch RP, Riegelman RK. Statistical First Aid: Interpretation of Health Research Data. Cambridge, MA: Blackwell Science; 1992.
24. Sexually Transmitted Disease Surveillance 2000 Supplement. Chlamydia Prevalence Monitoring Project. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2001.
25. San Francisco STD Prevention and Control Services. San Francisco Sexually Transmitted Disease Annual Summary, 2001. San Francisco: San Francisco Department of Public Health; 2002.