CAPTURE-RECAPTURE METHOD

TO ESTIMATE HIV PREVALENCE

IN ARGENTINE WOMEN

by

F. Anne Eades

A Thesis

Presented to the Department of Public Health and Preventive Medicine

and the Oregon Health Sciences University

in partial fulfillment of

the requirements for the degree of

Master of Public Health

Epidemiology and Biostatistics Track

April, 2001

School of Medicine

Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the MPH thesis of

F. Anne Eades

Has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

Firstly, I would like to thank the Public Health and Preventive Medicine department for giving me the opportunity to participate in the MPH program. This step in my education has been an exceptional experience. My professors have been not only good teachers, but also an inspiration. It has been the high point in my education. I thoroughly enjoyed my classes, the various projects, and my interesting and bright classmates.

I especially want to thank my committee: Dr. Jay Kravitz for his inspiration and undaunting faith that I could venture so far afield; Dr. Jodi Lapidus for her leadership, persistence, and kind nature, she has kept me on track and moving forward; and Dr. Mark Loveless for his expertise and enthusiasm.

This study would not have been possible without the help and guidance of Dr. Raul Mercer and the Programa Materno Infantil, Ministerio de Salud de la Provincia de Buenos Aires in Argentina. The people working there were wonderful. Dr. Mercer's commitment to the international exchange of students and information made it possible for me to join them for three months. He is a mentor and friend.

I want to thank my family, friends, and coworkers who encouraged me, and took care of the home front for three months so that I could leave on the trip of my lifetime.

Finally, to my eternally optimistic husband, Larry, who makes dreams come true.

ABSTRACT

HIV is a growing concern in Latin American countries, especially as it relates to women and children. While estimates of the number of AIDS cases in these countries exist, widely accepted estimates of HIV cases do not. In this study capture-recapture methods were used along with log linear modeling to estimate the number of HIV positive women in the Buenos Aires Province of Argentina. Three types of data sources were used: hospital records; a combined source of blood bank and public health laboratory records; and social service agencies' records. High quality data with unique identifiers were available from all of the sources, and all had data for a minimum of ten years.

For the ten year period from 1989 to 1998 there were 1598 entries identified for 1498 individual HIV positive women. There were crossovers for all data sources, with only one case crossing all three sources. The estimate of the total number of HIV positive women in the Buenos Aires Province is 9907 (95% CI 6543 – 15,516).

Capture-recapture methods offer an efficient and relatively quick method to gather data from existing information sources and form an accurate estimate of population size. It also provides a means to assess undercount and adjust rates accordingly. Public health officials can use this information to plan HIV prevention and treatment programs for women and children. Testing and treating women for HIVcan help infected women to lead longer, healthier lives, and make better decisions concerning pregnancy. Ultimately, this will help prevent many cases of HIV in Argentina's children.

BACKGROUND

Global HIV/AIDS Epidemic

While the international pandemic of human immunodeficiency virus (HIV) infection and acquired immunedeficiency disease syndrome (AIDS) is expanding rapidly, it is particularly affecting developing (underdeveloped) nations where most of the people of the world live. The Joint United Nations Programme on HIV/AIDS (UNAIDS) categorizes the world as: Eastern Europe and Central Asia, Western Europe, South and South-East Asia, Sub-Saharan Africa, North Africa and the Middle East, East Asia and Pacific Islands (excluding Japan), North America, Latin America (Mexico, Central America, South America), and the Caribbean. Of these regions, the continent of Africa is most affected by HIV/AIDS. It is home to 70% of the adults and 80% of the children living with HIV in the world, and contains three-quarters of the more than 20 million people worldwide who have died of AIDS since the epidemic began (UNAIDS, 2000). The developing countries have 90% of the HIV/AIDS cases and, by contrast, are benefiting little from the new developments in prevention and treatment (Mann, 1998). However, 90% of the money for care and prevention are being disproportionately spent in the industrialized nations of the world that include: North America, Western Europe, Japan, Australia and New Zealand.

Table 1. HIV/AIDS STATISTICS AND FEATURES BY WORLD REGION, THROUGH 2000

Region	Adults & children newly infected with HIV	Adults & children living with HIV/AIDS	Adult prevalence rate (*)	% of HIV-positive adults who are women	Main modes of transmission (#) for adults living with HIV/AIDS
Sub-Saharan Africa	3.8 million	25.3 million	8.8%	55%	Hetero
North Africa & Middle East	80 000	400 000	0.2%	40%	Hetero, IDU
South & South-East Asia	780 000	5.8 million	0.56%	35%	Hetero, IDU
East Asia & Pacific	130 000	640 000	0.07%	13%	IDU, Hetero, MSM
Latin America	150 000	1.4 million	0.5%	25%	MSM, IDU, Hetero
Caribbean	60 000	390 000	2.3%	35%	Hetero, MSM
Eastern Europe & Central Asia	250 000	700 000	0.35%	25%	IDU
Western Europe	30 000	540 000	0.24%	25%	MSM, IDU
North America	45 000	920 000	0.6%	20%	MSM, IDU, Hetero
Australia & New Zealand	500	15 000	0.13%	10%	MSM
TOTAL	5.3 million	36.1 million	1.1%	47%	

Data from AIDS Epidemic Update December, 2000, UNAIDS
*The proportion of adults (15 to 49 years of age) living with HIV/AIDS in 2000, using 2000 population numbers.
#Hetero (heterosexual transmission), IDU (transmission through injection drug use), MSM (sexual transmission among men who have sex with men), not in rank order.

The HIV/AIDS epidemic varies by region of the world (Table 1). In all parts of the world except sub-Saharan Africa, there are more men infected with HIV and dying of AIDS than women. Men's cultural beliefs about masculinity usually contribute to their high-risk behaviors and increased risk of HIV exposure. Male risk behavior also contributes to HIV infections in women, who may have less authority to determine the circumstances of sexual encounters (UNAIDS, 2000). Transmission of HIV from men to women is more efficient than the reverse direction. Factors that favor transmission of HIV to women include: semen has a higher concentration of HIV than vaginal secretions, the female anatomy has a large surface area where infected semen can remain for a prolonged time, women with sexually transmitted diseases have more entry points for HIV, presence of inflammation in the vagina may provide target cells for HIV and increase risk of transmission.

The HIV epidemic has followed similar patterns of transmission from country to country. The epidemic usually starts in, and spreads, through the male population. Transmission is enhanced by unsafe injection drug use practices. The exception is Africa, where injection drug use is relatively uncommon. Early phases of sexual HIV transmission are aided by inadequate condom use, much of it through commercial sex and some through sex between men (MSM). When HIV spreads mainly within a small subgroup of the population, such as men who have sex with men, as it did in North America, the number of potential exposures is limited. However, even among MSM, bisexuality and drug use provide bridges to the general population. As the epidemic matures, HIV is transmitted through sex between men and women. This represents a risk to a far larger proportion of the whole population. Increasingly, women become infected and begin to equalize or overtake males. The involvement of women in the epidemic ultimately leads to the addition of vertical transmission of HIV from mother to infant. Vertical transmission occurs when the virus moves from the mother to the child during pregnancy, delivery, and/or through nursing. By studying the HIV prevalence in a population, one can determine what stage the epidemic is in. For example, as seen in Table 1, Africa's main mode of transmission is heterosexual contact indicating a maturing epidemic. Eastern Europe, on the other hand, has injection drug use as the main mode of transmission, indicating the HIV epidemic is in the early stages there.

Social and behavioral factors contributing to the sexual transmission of HIV include (UNAIDS, 2000):

- Little or no condom use
- Large proportion of the adult population with multiple partners
- Overlapping (as opposed to serial) sexual partnerships. [1]
- Large sexual networks (often seen in individuals who move back and forth between home and a far-off workplace)
- "Age mixing", typically between older men and young women or girls
- Women's economic dependence on marriage or prostitution, (denying them of control over the circumstances or safety of sex.)

Biological factors that contribute to HIV transmission include (UNAIDS, 2000):

- High rates of sexually transmitted infections, especially those causing genital ulcers
- Low rates of male circumcision
- High viral load – HIV levels in the bloodstream are typically highest when a person is first infected and again in the late stages of illness.

Programs around the world work to reduce risk factors for HIV exposure. Sentinel groups such as pregnant women, blood donors, and injection drug users are surveyed for HIV. Surveillance helps to describe risk factors for targeted prevention efforts and to determine not only the prevalence of HIV disease, but also the maturity of the epidemic for the area.

HIV/AIDS in Latin America and the Caribbean

There were an estimated 1.4 million adults and children in the Latin American and Caribbean region living with HIV or AIDS at the end of 2000, as compared with 1.3 million at the end of 1999 (UNAIDS, 2000). The Pan American Health Organization (PAHO) divides South America into Brazil, the Andean Area (Bolivia, Colombia, Ecuador, Peru, and Venezuela), and the Southern Cone (Argentina, Chile, Paraguay, and Uruguay). Areas vary with regard to the predominant transmission patterns. Principal

---

[1] Individuals are highly infectious when they first acquire HIV and thus more likely to infect any concurrent partners.

transmission patterns for the Andean Area are: homosexual/bisexual contact 42.3%, heterosexual contact 40.8%, and injection drug use <1% of HIV transmission. In the Southern Cone HIV transmission patterns are: homosexual/bisexual contact 32.3%, heterosexual contact 22.4%, and injection drug use accounts for 33.4%. Brazil's HIV transmission patterns are similar to the Southern Cone with 28.7% homosexual/bisexual contact, 23.6% heterosexual contact, and 19.3% injection drug use (PAHO Report, 2000).

Central America reported heterosexual transmission for 72.8% of its AIDS cases, and 11.8% from homosexual/bisexual transmission. Mexico's pattern of transmission was recorded as 35.1% due to homosexual/bisexual contact, and 25% heterosexual contact (PAHO Report, 2000).

The Caribbean region is experiencing one of the worst epidemics outside Africa with an estimated adult HIV seroprevalence of 2.3%. Sub-Saharan Africa's seroprevalence is estimated at 8.8%. The percentages vary between the Caribbean countries, from 0.02% in Cuba to 12% in Haiti. In Guyana, blood donors have a seroprevalence of 3.2%, while urban sex workers have a seroprevalence of 46% (Editorial, Lancet, 2000).

HIV/AIDS in Women and Children

As the HIV epidemic matures in an area, heterosexual contact becomes the primary form of transmission. When this happens, HIV disease becomes, increasingly, a health problem affecting women. The proportion of women with HIV/AIDS relative to the number of men begins to equalize. Women are often infected as a result of a partner's high risk sexual behavior rather than their own. Heterosexual contact is the major risk factor commonly identified for women as opposed to the risks identified in immature epidemics such as injection drug use and commercial sex work. In regions with mature epidemics, such as Sub-Saharan Africa and the Caribbean, women have higher infection rates than men – a reflection of the greater efficiency of male-to-female HIV transmission, as well as women's increased vulnerability to sexually transmitted diseases (STDs) in general (Tarantola, 1996). A worrying trend is the steadily increasing prevalence among women and children in many parts of the region with the male to female ratio approaching 1 to 1 in Haiti and the Dominican Republic (Editorial, Lancet, 2000). In Sub-Saharan Africa, the proportion of HIV-infected women is greater than the proportion of HIV infected men. The male to female ratio is 1:1.1 (UNAIDS, 2000).

Increases in HIV-infected women lead to increased vertical transmission, from mother-to-infant. HIV may be transmitted to an infant through the mother's milk, by contact with maternal blood and/or genital secretions during labor, and by virus crossing the placenta (Douglas, 1992). Since most cases of AIDS in children result from vertical transmission, HIV infection in childbearing women is the major factor in determining the trend in childhood AIDS incidence. The risk of transmission of HIV from mother-to-infant is between 13% - 40% (Havlir, 1996). At least 50% of maternal-fetal transmission occurs late in pregnancy and during delivery (Green, 1992).

The level of HIV in the mother's bloodstream, or viral load, increases the likelihood of transmission. Therapeutic strategies that aim to reduce viral burden in the mother at delivery or to treat the neonate after exposure are effective in reducing HIV disease in infants born to infected mothers. Havlir estimates that transmission can be reduced by two thirds, if the HIV infected mother is identified early and treated (Havlir, 1996).

HIV-infected mothers in developing countries often learn of their HIV status because they have given birth to an infected infant (Lallemant, 1994). Interventions to prevent mother-to-infant transmission of HIV, including recent breakthroughs in antiretroviral therapy, such as zidovudine, offer immediate opportunities to save children's lives. The UNAIDS/World Health Organization recommends that pregnant women, and those contemplating pregnancy, should have access to voluntary HIV testing and counseling. All pregnant women should have access to antenatal, delivery and post-partum care, and to a skilled attendant at birth to minimize maternal and infant risks, assure proper administration of zidovudine, and to decrease the infant's exposure at the time of labor and delivery. For the zidovudine regimen to be effective, at least one antenatal visit with follow up is needed before 36 weeks, and preferably before 34 weeks, of gestation. For HIV positive mothers, skilled care during delivery is needed to assure the proper administration of antiretroviral therapy, zidovudine, during labor and delivery (UNAIDS/WHO, 1998).

In 1998, a CDC-sponsored trial in Thailand demonstrated that there is a zidovudine regimen that is effective and feasible in developing countries. The short regimen involves the administration of zidovudine to mothers during the last four weeks of pregnancy and during delivery. It has been shown to reduce mother-to-infant transmission by half among women who do not breastfeed. In many developing countries this intervention is cost-effective when compared to other public health interventions in maternal and child health (UNAIDS/WHO, 1998).

HIV/AIDS and Women and Children in Argentina

Argentina is located in the southern cone of South America. It is the eighth largest country in the world with 1.1 million square miles. The islands and Antarctic land together cover an additional 480,000 square miles. It is the second largest country in South America behind Brazil. Argentina is 2,170 miles long and 868 miles across at its widest point.

The Buenos Aires Province holds 38.6% of Argentina's population of 36,577,000 (UNAIDS, 2000), and 78% of the AIDS cases among Argentine women. The principal cities are La Plata, which is the capital of the province, and Buenos Aires, which is the capital of the country. The city of Buenos Aires contains 63% of the population of the province and 24% of the country. About 86% of the Argentine population live in urban areas (Bernhardson, 1996). The population of Argentina aged between 15 – 49 years is approximately 18.2 million, about half of the total population. The numbers of men and women in that age group are approximately equal (UNAIDS, 2000).

Argentina is a country of European immigrants: Italians, Basques, Welsh, English, Ukrainians, Germans, and others. The Buenos Aires' Jewish community of 400,000 is the eighth largest in the world. Argentina's indigenous population is estimated to be 100,000 people and, for the most part, inhabit northern Patagonia (Bernhardson, 1996).

In Argentina, the first case of AIDS in a woman was reported in 1987 (five years later than the first male case). The number of cases among women has continued to increase from one in 1987 to a high of 558 AIDS cases in 1996. The cumulative number of AIDS cases in women through 1998 was 2976.

In 1987, in Buenos Aires, 36 percent of injection drug users tested were HIV positive, by 1995, over 90 percent of injection drug users tested were HIV positive (UNAIDS Factsheet, 2000). In 1994, the most common risk factor among men with AIDS, was injection drug use, while the most common risk factor among women with AIDS was heterosexual contact (representing 37% of total reported cases in women with AIDS) (Raggi, 1992, Betts, 1996).

Few seroprevalence studies have been done that address the issue of heterosexual HIV transmission in Argentina. A seroprevalence study of a relatively small sample of 2311 pregnant Argentine women was done in 1988, early in the epidemic, and observed a small number of HIV+ cases. It yielded a rate of HIV prevalence of 0.82%. Seventeen of the nineteen women identified in this study as HIV positive were injection drug users. The remaining two women were infected through heterosexual exposure, one to an injecting drug user, and the other to a bisexual male. This pattern of transmission is common in the early phases of the epidemic. Another study was conducted six years further into the epidemic, in 1994 –1995. Women tested were living in Buenos Aires, and receiving prenatal care. A total of 482 pregnant women were tested. The prevalence was 3.5%, significantly higher than the 1988 study (Betts, 1996). Testing of 44,140 pregnant women in the southern part of Buenos Aires city in 1998 yielded an HIV median prevalence of 0.29%, with a maximum prevalence of 2.9%. Prevalence varied depending on the subpopulations tested (Boletin, 1999).

The HIV epidemic in Argentina shows signs of maturing. It has moved from men who practice high-risk behaviors to women and children. The ratios of men to women with AIDS have shifted over the last decade. In Argentina in 1989 the ratio of AIDS men to women was 8.5:1. By 1998 the gap had closed to 2.6:1 (Boletin, 1999). In Argentina, 48% of women with AIDS contracted it via heterosexual activity, 33% via exposure to blood, and 16% from an HIV+ mother. Of the women in Argentina who have AIDS, 78% are between the ages of 15 and 44 years, and 78% of the Argentine women with AIDS live in the Buenos Aires province. The cities of Argentina house 93% of the total cases of AIDS in the country. The largest concentration of AIDS cases in Argentina are in Buenos Aires (4766), suburban Buenos Aires (4783), Rosario (830), Cordoba (292), La Plata (251), Mar del Plata (186), and Mendoza (111). The other cities by 1999 had contributed less than 100 cases (Boletin, 1999).

With the exception of a few seroprevalence surveys, and monitoring of certain sentinel populations, most of the available information on transmission of HIV in Argentina comes from data on AIDS cases. HIV cases are not monitored. Examining AIDS data in lieu of HIV data can only indicate transmission patterns well into the past due to the long incubation period from the time of infection with HIV (up to ten years). While information from other countries that are farther along in the evolution of their AIDS epidemic may be helpful, it only points the way for further investigation in Argentina. Other methods that utilize exiting information are needed to effectively estimate the number of HIV+ women in the country.

Despite the WHO recommendation for testing of pregnant women, women in Argentina are not routinely tested for HIV. Pregnant women are tested if there is knowledge of risk factors for HIV exposure such as: injection drug use, commercial sex work, or having sexual contact with a man with known risk factors. However, the principal risk factor for women remains unprotected heterosexual activity which is not a risk factor that would result in testing (Boletin, 1999). Development of an accurate estimate of women who are HIV positive could encourage and direct support to prevention and treatment programs of HIV in women.

Capture-Recapture Method of Estimation

The capture-recapture method of estimation is used to estimate an unknown population size, such as the number of HIV+ women in a given area of Argentina. The methodology utilizes multiple lists of individuals known to be in the population, and determines the overlap of such lists. Then, the unknown population size, is estimated using log-linear modeling techniques. An appealing feature of capture-recapture is that it does not require complete information from any source. Multiple incomplete lists can provide the necessary information.

The methodology has a long history of use in demography, from estimating the population of France by Laplace in 1786, determining birth and death rates in India, to uses in the1990 US Census. The implementation and acceptance of this approach in biology and health care began with the formulation by Sekar and Deming in the late 1940s (Sekar, 1949), followed by classical mathematical developments by Wittes (Wittes, 1974) and Seber (Seber, 1982). These methods have considerable potential in epidemiology for monitoring diseases. Even exhaustive studies can benefit from capture-recapture to estimate, or adjust for, missing cases. Surveys that report prevalence data without adjustment for, or data on, source intersection in essence make an estimate of missing cases of zero. Such studies essentially are providing only a lower boundary of prevalence (Hook, 1992).

Despite the epidemiologic potential, the greatest application of capture-recapture has been in the monitoring of wildlife, where it has become the standard for estimating populations of birds, fish, and insects. To estimate the size of a population of fish, one would go out, catch fish, tag them, and then release them. On subsequent days one would net fish again and note the number of tagged fish in the catch. An estimate of the population size can then be calculated based on the number of marked fish captured in the second sample (McCarty 1992, 1993; LaPorte, 1994). The name capture-recapture is based on these applications of the method.

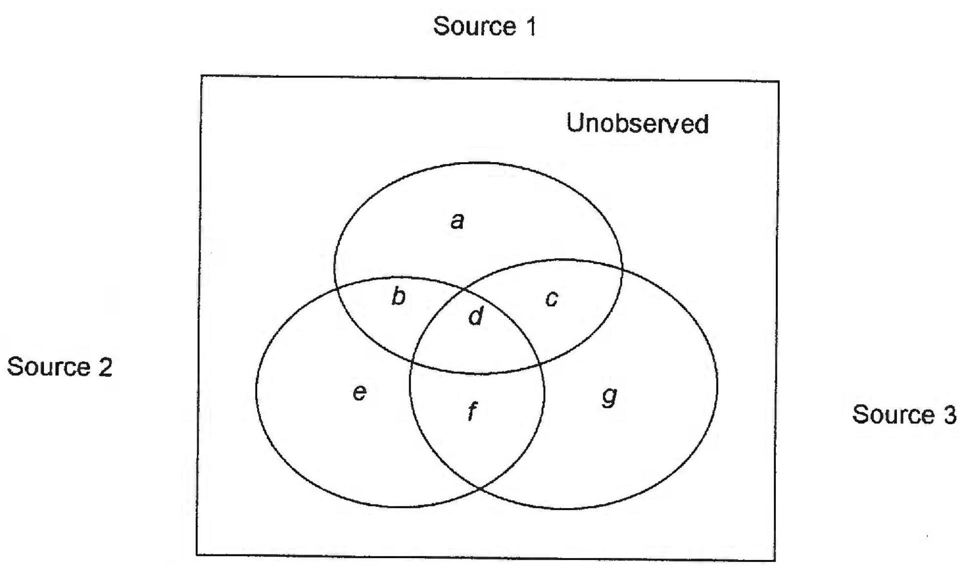
Traditional public health approaches to counting of diseases – surveillance, registries, and death certificates do not always meet the needs for the monitoring and forecasting of disease occurrence. These systems are too inaccurate (surveillance), or too costly (population based registries), or too late (death certificates) (LaPorte, 1992). In the United States, the cost of monitoring of all serious diseases with incidence registries is estimated in excess of five billion dollars (International Working Group, 1995). Developing countries are faced with both serious infectious diseases and increasing chronic diseases, but the necessary resources and infrastructure to monitor them are often non-existent. Capture-recapture may be the only practical technique to assess disease frequency in developing countries (International Work Group, 1995).

Capture-recapture is a desirable alternative to traditional counting methods because it utilizes existing information. It is less expensive to manage, and more timely. Capture-recapture can be used to estimate incidence rates and prevalence that are more accurate. All over the world, serious diseases are documented on various lists by those who test for them, provide care for those affected, or provide other kinds of support.

Incomplete lists of individuals are both readily available and inexpensive to use. For any human condition that is important to measure and monitor, there are many sources from which the individuals may be identified. The more serious the condition, the greater the likelihood that it is listed. This information can be found in doctor's offices, clinics, pharmacies, hospitals, and agencies. By using existing information in various databases, including the use of duplicates between samples, capture-recapture can estimate the population of interest (Int. Working Group, 1995).

The simplest capture-recapture method, the two-sample model, works as follows. The first sample is collected, and each subject identified with a unique identifier. A second sample is collected and compared to the first by the same unique identifier. Using the numbers of individuals "caught" in both samples (the recaptures) and the numbers caught in just one of the two samples, it is possible to estimate the number not in either sample. Adding the numbers caught plus the estimate of those not caught, it is possible to obtain an estimate of the total population size. By taking more than two samples, one can utilize the information from the multiple recaptures to estimate the population size.

Figure 1. CAPTURE-RECAPTURE CONCEPTUAL MODEL



a = number of cases that were from source 1 but not from sources 2 or 3.
e = number of cases that were from source 2 but not from either source 1 or 3.
g = number of cases that were from source 3 but not from either source 1 or 2.
b = number of cases that were identified by both sources 1 and 2 but not by source 3.
c = number of cases that were identified by both sources 1 and 3 but not by source 2.
f = number of cases that were identified by both sources 2 and 3 but not by source 1.
d = number of cases that were identified by all three data sources.

Figure 1 illustrates the conceptual framework for a 3-source capture-recapture analysis. Each circle represents data from a source. The interlocking areas represent recaptured data between sources. The area where the circles meet is data that was identified by two (b,c,f) or three (d) sources. The area outside of the circles and within the box represent the unobserved data, or the remaining unknown population. Estimating the unobserved region is the objective of capture-recapture.

Table 2. CAPTURE-RECAPTURE ASSUMPTIONS

Number	Assumption	Terms
1	There are no additions or deletions to the population during the investigation	Closed population
2	There is no loss of tags, individuals can be identified from capture to recapture	Identifiability
3	For each sample, each individual has the same chance of being in the sample	Homogeneity
4	Capture in the first sample does not affect capture in the second sample	Independence

Table 2 outlines the assumptions required for the capture-recapture estimate to be valid as taken from the capture-recapture web site (Chang, 1999). The first assumption that the population remains unchanged during the study period refers to a closed population. If individuals migrate into an area and/or are lost because of death or out-migration in the study interval, then the population is considered open and violates the assumption of constant catchability. By migrating in or out, such as moving into a different province or country, individuals change their chances of appearing on lists equally, or demonstrate variable catchability. Almost all populations available to epidemiologists, violate constant catchability. To model open populations, it would be necessary to have additional external information on exit probabilities from various sites. This information is usually unavailable. Most populations studied in epidemiology with capture-recapture are semi-closed, closed to "in-migration" but not to "out-migration" or death (Hook, 1995).

Loss of tags is a problem that generally occurs in wildlife application of capture-recapture. However, in human populations, identification errors can occur which will effect matching across different sources. People can use different identifying information at different sites, plus errors in recording can be made. For studies involving HIV, confidentiality can be an issue. In Argentina, the data are coded according to an international coding convention with the individuals' initials in a particular order, and with their date of birth. Addresses can be limited to cities, only. Careful attention to data and data entry can minimize much of the potential for error.

The most difficult assumption to accommodate is number three in Table 2, that individuals have equal chance of appearing on a list, or that the population is homogeneous. Lack of homogeneity, or the presence of heterogeneity, within the sources is difficult to determine, and can effect overall estimates of the population. Subgroups within a source can vary in their catchability due to differences in age, sex, severity of disease, geographic location, or by other social or behavioral factors.

Using more sources of data across greater geographical or social groups can mitigate heterogeneity. For example, investigators of opiate users in Barcelona observed heterogeneity in different age, sex, and residence area subgroups. This caused population estimates to differ widely according to the model chosen (Domingo-Salvany, 1998). When more sources are available, pooling or collapsing of sources can be useful. The results of multiple different types of combination of sources may be particularly effective in overcoming the problem of subgroup variation.

The other difficult assumption to accommodate is independence between sources, number four in Table 2. In practice, truly independent sources rarely occur. For instance, physicians tend to refer patients to specific hospitals, laboratories, and certain specialists. Positive dependence between sources occurs when individuals are more likely to be identified by particular source versus any source. Negative dependence occurs where individuals may be systematically excluded from a selected source. The direction of the dependence directly effects the precision of the estimations. Positive dependence deflates estimates, while negative dependence increases estimates (McCarty, 1993). If the directions of the dependencies are known, or highly suspected, then biased estimates can still be useful as indicators of upper or lower boundary limits (Hook, 1995). There is a trade-off between stratification and the modeling of dependence, since increasing stratification reduces the sizes of the counts in individual strata and this decreases the probability of detecting important interactions within strata (Hook, 1993, International Working Group, 1994, Domingo-Salvany, 1998).

If dependencies or interactions between two sources are suspected, it is recommended that three or more sources be employed to control for potential dependencies prior to correcting ascertainment for incidence or prevalence. Ismail found three sources with identifiers for the first and last name and date of birth were adequate to model for dependence between data sets. Using more than three lists did not substantially alter the absolute value or confidence of enumeration. Multiple lists, if available, should be condensed into three lists for use in capture-recapture calculations (Ismail, 2000).

Log Linear Modeling in Capture-Recapture

Table 3. INCOMPLETE MULTI-WAY CONTINGENCY TABLE

		Source 1			
		Yes		No	
		Source 2		Source 2	
		Yes	No	Yes	No
Source 3	Yes	a	b	e	f
	No	c	d	g	x=0

Log-linear models were first proposed by Fienberg to offer a technique to address dependency between sources (Fienberg, 1972). Using this technique, data can be placed in an incomplete 2^k contingency table, where k represents the number of sources selected for case ascertainment. Table 3 demonstrates an incomplete multi-way contingency table. This table displays the recapture history of all observed cases in the population. There is one missing cell (x) that corresponds to the number of unascertained cases in the population. Log-linear models are fitted to this contingency table which uses the observed cells to predict the missing cell. By predicting the value of the missing cell, estimates of the size of the population of interest may be obtained. Source dependencies may be modeled using interaction terms that are added to the simplest model (independent sources) until no significant improvement of fit occurs (McCarty 1993, 1995).

Model selection is an integral part of the estimation procedure. In the early days of capture-recapture there were few models to choose from, and the most appropriate model of those available was easy to identify. Today there are a large number of potential models designed to handle dependence between lists and heterogeneity between individuals. The development of software to handle modeling has made the process much easier. With three lists there are eight possible combinations of lists in which individuals are observed or unobserved. In the general model there are eight parameters, the common u – the logarithm of the number expected to be in all lists, three "main effects" u_A, u_B, u_C – log odds against appearing in each list for individuals who appear in the others, three "two-factor interactions" u_{AB}, u_{AC}, u_{BC} – log odds ratios between pairs of lists for individuals who appear in the other, and a three-factor interaction u_{ABC}.. The likelihood ratio statistic, G^2 can be used to choose an estimate associated with the single model that is judged optimal. The lower the value of G^2, the better the fit of the model (International Working Group, 1995).

An information criterion that is a function of its likelihood ratio statistic can be used to determine the optimal model. The Akaike Information Criterion or AIC, and the Bayesian Information Criterion (BIC) are commonly used for capture-recapture models. The optimal model is the one with the lowest value of the information criterion. For the saturated model, because both G^2 and degrees of freedom are zero, AIC and BIC are also zero. Thus, by these criteria, only models with negative values of the criteria are preferable to the model with all of the interactions included (saturated model). If there are more than about 46 observed cases, then the BIC picks simpler models. The AIC is more widely used. There may be little difference in the performance of the AIC versus the BIC, however, they sometimes will result in different "best model" choices (Hook, 1995).

Most authors will admit that the model selection procedure does not identify a selected model that is true and the only one worth considering. There may be statistical evidence that one of the acceptable models is marginally better than the others. However, it is important to realize that one model and its estimate, even with a narrow theoretical confidence interval, do not represent the absolute truth. Instead, there can be a range of acceptable models with their estimates providing a range of values that contribute to our knowledge of a population. The BIC weighted average estimate, proposed by Draper, takes an average of the estimates from all models' fit, and weights each estimate by the BIC associated with its model. By weighting the BIC from a particular model, estimates from models that fit well will be weighted more heavily than those from models that do not fit as well. It has been found by some investigators to be preferable to either the AIC or the BIC alone, particularly in situations where more than one model have approximately equal BICs but different parameter estimates (Hook, 1995).

Capture-Recapture Examples in Public Health

Capture-recapture has been applied in many categories of disease monitoring including birth defects (Egeland, 1995), cancer (Ballivet, 2000), drug use (Mastro, 1994;Domingo-Salvany, 1995, 1998), infectious diseases (Sanghavi, 1998; Modesitt, 1990; Ackman, 1996), injuries (LaPorte, 1995), insulin dependent diabetes (LaPorte, 1993; Bruno, 1994), and in other studies of hemophilia (Hewitt, 1970), myocardial infarctions (LaPorte, 1992), size of homeless populations (Fisher, 1994), and the number of children dependent on medical support (Palfrey, 1991). It has also been used to evaluate the effectiveness of surveillance systems for monitoring abortion mortality (Cates, 1978), infections among hospitalized patients (Lewis, 1969), Rett's syndrome (Kozinetz, 1993), and vaccine-associated paralytic poliomyelitis (Prevots, 1994). Even when exhaustive case finding is used for a study, capture-recapture can evaluate the effectiveness of the surveillance system by correcting for ascertainment (Hook, 1992).

The following six examples of capture-recapture studies demonstrate the usefulness of the method for a variety of settings and difficult-to-capture populations.

A two-sample capture-recapture study in Bangkok, Thailand was used to estimate the number of HIV-infected injection drug users (Mastro, 1994). The study yielded an estimate of approximately 12,000 HIV-infected injection drug users in Bangkok in 1991. The authors concluded that capture-recapture was useful in estimating difficult-to-reach populations, including injection drug users. Estimates indicated that in the future the number of cases of AIDS and other HIV-related diseases would increase and place new demands on existing health care facilities. More accurate estimates would allow for planning of health services for this population (Mastro, 1994).

In Oregon, USA, capture-recapture was used by the Oregon Health Division to determine the completeness of reporting in a passive versus an active AIDS surveillance program. Historically, the passive surveillance system had a 64% completeness of reporting. An active surveillance system was put into place, and had a completeness of reporting of 99%. Through the use of capture-recapture, several data sources were identified as targets for enhanced surveillance activities. This would increase the efficiency of staff to trend and plan for the epidemic. The authors concluded that as the epidemic increases, an active surveillance system would best estimate the population (Modesitt, 1990).

Capture-recapture was used in a study in New Caledonia in the South Pacific to determine the optimum combination of sources for a thyroid cancer surveillance system. The preferred combination included four sources. The assessment of dependence, based on simple criteria, was used to choose the best combination of sources for the surveillance system. Capture-recapture methods can be used to assess a surveillance system and increase its productivity (Ballivet, 2000).

In London, researchers used the multiple sample capture-recapture method to estimate the number of homeless and homeless mentally ill in northeast Westminster, (Fisher, 1994). Lists were developed compiling routinely collected data from local agencies. The study estimated the population of unobserved homeless (5000) to be more than twice the size of the observed population (1640). The disparity between the observed and unobserved population was attributed to the fact that homeless people with mental health problems were disproportionately visible to the sources that were used. Without correction for ascertainment, the net effect would be an overestimate of the homeless, and an underestimate of the homeless mentally ill. The authors concluded that capture-recapture can overcome problems of ascertainment in the estimation of homeless populations. Capture-recapture techniques are useful to provide a basis for including the homeless in capitation calculations for allocating funding in health services (Fisher, 1994).

In Barcelona, two capture-recapture studies were conducted by the same investigators in 1989, and in 1993, to estimate the population of opiate addicts

(Domingo-Salvany, 1995,1998). Multiple sources were used to identify addicted individuals. The authors determined that the capture-recapture method underestimated the number of opiate addicts because some subgroups of addicts (women, and older men) were consistently under represented on the available lists. Adequate sample size, and adjustment for heterogeneity improves the accuracy of estimates. Despite problems with their final estimate, the authors felt that capture-recapture was a good method to estimate the prevalence of drug addiction, and may be the only way to achieve ascertainment-corrected estimates on this elusive population. The authors emphasized the necessity for accurate estimates of drug addiction to plan and evaluate public health interventions for this group.

A capture-recapture study was performed in a Peruvian shantytown of 34,000 to estimate the incidence of TB in a hyperendemic setting (Sanghavi, 1998). The authors concluded that there were clusters of hyperendemic TB that were not reported accurately by the health ministry because data from different zones with very different incidence rates were pooled. The government reported average annual incidence as 134/100,000. The capture-recapture method yielded 364/100,000 population, an increase of 170%. The authors recommended that broadly targeted measures for TB control be used in poverty-stricken, hyperendemic shantytowns. The authors also recommended that a capture-recapture study be performed periodically to establish the accuracy of Ministry reporting.

Each study used capture-recapture to provide an accurate estimate of a difficult-to-capture population. In two of the six studies, the Oregon study and the thyroid cancer study, capture-recapture was also used to determine the best sources for an efficient surveillance system. Ultimately, the information from estimates of population size can help to direct and target resources for health services to the affected groups.

GOALS AND AIMS

The goal of this study was to provide useful information for Argentina's public health decision makers in the fight against HIV/AIDS. The capture-recapture method was applied to achieve estimates of the number of HIV+ women in the Buenos Aires Province of Argentina using the capture-recapture method. This was accomplished without a large staff, a lot of time, or a public health infrastructure. The information from the study can be used along with other information about the AIDS epidemic in Argentina to plan strategies and interventions that will especially help the women and children of that country.

Secondarily, in the process of doing the study, cultural obstacles were overcome, ideas were exchanged, and international connections were made. For me, personally, I had the opportunity to visit a distant part of the world, and make a small contribution towards its well-being.

MATERIALS AND METHODS

This study was performed in March, April, and May of 1999 when the author visited Argentina. The target population was HIV+ women twelve years or older living in the province. Dr. Raul Mercer, director of the Maternal and Child Health Program (Programa Materno y Infantil, Minsterio de Salud de la Provincia de Buenos Aires) in the city of La Plata was my contact person and facilitated the study. Dr. Mercer was interested in the study because the AIDS epidemic in Argentina was growing and increasingly affecting women and children. His wife worked at the large children's hospital in La Plata where more and more children were being admitted with AIDS. It was a public health concern as well as a personal one. A more accurate estimate of the prevalence of HIV in women could help in planning for future services for women and children.

Dr. Mercer had previously been in contact with Dr. Ron LaPorte of the University of Pittsburgh. Dr. LaPorte is a leader in the use of capture-recapture for public health studies. Through e-mail we discussed the study with him. I also reviewed available information through searches on the Internet and at the scientific library in La Plata.

Dr. Mercer and his staff sent a letter of introduction to potential sites for the study in advance of my visit. After I arrived, Dr. Mercer made additional contacts by telephone to study sites and other resources such as the scientific library (Federation Bioquimica Biblioteca) and the laboratory quality-testing agency (Federacion Bioquimica Control de Calidad). Specific dates and times for meetings were set up over the telephone with hospital and laboratory officials. I then would visit the site, explain the study, and elicit their participation. Because I was not fluent in Spanish, it was sometimes difficult going to the sites alone. The people there were very patient with my lack of command of the language.

There were a total of seven facilities and agencies used as sources of data for this study as shown in Table 4. They included hospitals and laboratories, both public and private, which do HIV testing for the Buenos Aires Province. There also were two public assistance agencies that financially support people living with HIV.

HIV testing was performed in the laboratories and hospitals that were staffed by licensed and trained professionals. The methods and equipment used in the laboratories were up-to-date and current with today's technology. A variety of primary screening tests are used, such as latex agglutination, enzyme immunoassay, immunofluorescence, electroimmunofluorescence, and radioimmunoassay. Positive screening tests are confirmed by the Western Blot test. The laboratory quality-testing agency, Federation Bioquimica Control de Calidad, in La Plata, confirmed proper standards for testing and quality control methods.

Table 4. FACILITIES AND AGENCIES USED AS DATA SOURCES

Category	Name	Features
Hospital	Hospital General Manuel Belgrano, San Martin (BGO)	Inner city community hospital, associated with a university, low-income area
Hospital	Florencia Varela Hospital Mi Pueblo (MP)	Community hospital, lower middle-class neighborhood
Hospital	Hospital San Juan de Dios (SJD)	Large regional hospital, specializes in HIV/AIDS, receives referrals from other hospitals
Laboratory	Laboratorio Central de Salud Publica (PHL)	Large, regional, public health testing center
Laboratory	Instituto de Hemopterapia (BB)	Regional blood bank, located in middle-class neighborhood, one of several in the province
Agencies	Instituto de Obra Medico Asistencial (IOMA)	Government-supported public assistance program, similar to social security
Agencies	Programa Atencion Medica Integral (PAMI)	Assistance program that benefits public employees

At some sites, laboratory records were in the process of being computerized, but for the most part, I used handwritten logs of HIV test results. When I was unsure of the writing, one of the technicians or secretaries would help clarify. The two social service agencies and the blood bank provided printouts of positive HIV cases from which I chose cases that met my criteria. I was the only person that collected and recorded data for the study.

I recorded data from handwritten lists directly into Excel spreadsheets on a laptop computer. The data items for each subject included: data source, subject initials, date of birth, and year of HIV test. A separate Excel worksheet was created for each source's data. All of the sources were very consistent in record keeping, and had standard information for the time period, 1989-1998. All of the logs used an international coding convention to identify the individuals. The two identifiers used were the individuals' initials, and the date of birth. Specific individuals could be identified on multiple lists from different sources.

The records also included the city the person was from, and the gender of the person. Some sources included risk factors for HIV and if the HIV positive case had died. However, this information was not consistently available and I did not record it. All of the sources had the year of the positive test. One of the hospitals did not have all of the dates of the positive tests. I recorded this information for those that were available.

A final aggregate worksheet was created for matching cases. Cases were matched two ways: first by initials, and secondly by date of birth. As an edit check, cases were also matched by date of birth and then initials. No additional matches were found.

Ethical issues were minimal since all data came from already existing lists. Records of persons who were HIV+ were coded by an international convention using initials and dates of birth. All information remained anonymous and confidential.

Permission to use the data was acquired from each participating facility. I was not required to sign any confidentiality forms. However, I did explain how and where the data would be recorded and used. Individual patient consents were not required. All results are reported as aggregate information.

DATA ANALYSIS

Based on literature reviews, I made the decision to combine the seven sources into three categories, hospitals, laboratories, and social service agencies for capture-recapture analysis (Hook, 1995; Domingo-Salvany, 1998; Fisher, 1994). Table 4 demonstrates the three categories used with the blood bank grouped with the laboratories. Three sources of data lead to eight possible log-linear models (refer to Figure 1): one main effects model, three single interaction models, three two-interaction models, and one saturated model. Log-linear modeling was used, and a structural zero was incorporated into the model since all possible cases of HIV could not be observed. Analysis was performed using SPSS, version 10.

All eight possible models were examined. Significant interactions, or source dependencies, between the laboratories and the hospitals and the laboratories and the social service agencies were identified. Both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used to determine the best fitting model. Confidence intervals were calculated for the population estimates. The optimal model is the model with the lowest value of the Akaike Information Criterion (AIC). The "weighted BIC" weights the estimate from each model. Hook suggests that it may perform better overall than approaches that select a single model using either AIC or BIC, (Hook, 1995). The final estimate uses this method.

RESULTS

There were 1598 individual data entries for 1498 persons across the seven data sources. Twenty-six cases from the Belgrano hospital were missing the date of the positive HIV test. This did not effect the capture-recapture analysis.

Table 5. CASES OF HIV POSITIVE WOMEN SORTED BY YEAR OF TEST AND BY TESTING SOURCE

SOURCES

Test Year	BB	BGO	IOMA	MP	PAMI	PHL	SJD	Totals
1986			1					1
1987			2					2
1988			5				16	21
1989	1		10		2	101	16	130
1990	1		11		5	128	7	152
1991	1		6		11	75	13	106
1992			23	3	13	128	83	250
1993	1	3	25	2	15	105	29	180
1994	5	7	22	3	14	28	24	103
1995	1	6	26	5	11	36	21	106
1996	4	4	45	3	21	23	38	138
1997	4	7	55	7	13	38	58	182
1998	6	13	54	15	18	61	34	201
Totals	24	40	285	38	123	723	339	1572

BB-Blood Bank, BGO-Belgrano Hospital, IOMA-social service agency, MP-Mi Pueblo Hospital, PAMI-social service agency, PHL-public health laboratory, SJD-San Juan de Dios Hospital

Table 5 shows the dispersion of cases across the different sources for the different test years. The public health laboratory (PHL), the HIV regional treatment hospital (SJD), and the public service agency, (IOMA), were the largest contributors of cases.

Table 6. HIV CASES SORTED BY AGE CATEGORIES AND BY TESTING SOURCE

	Age Categories							
Sources	12-19	20-29	30-39	40-49	50+	Unknown	**Totals**	% of Total
BB	1	8	9	4	2		24	1.50%
BGO	5	19	13	3		26	66	4%
IOMA	20	140	91	27	7		285	18%
MP	5	21	10	2			38	2%
PAMI	9	54	32	11	17		123	8%
PHL	118	453	118	28	6		723	45%
SJD	65	175	75	18	6		339	21%
Totals	223	870	348	93	38	26	1598	
% of Total	14%	54%	22%	6%	2%	2%		

Sources: BB-Blood Bank, BGO-Belgrano Hospital, IOMA-public service agency, MP-Mi Pueblo Hospital, PAMI-public service agency, PHL-public health laboratory, SJD-San Juan de Dios Hospital

Table 6 shows the distribution of ages across the different sources. The greatest numbers of cases are in the younger age groups, with the largest proportion (54%) in the 20-29 category. This is consistent with published data for AIDS cases among Argentine women. Women tend to be younger when diagnosed with AIDS than men. Published AIDS data indicate that the average age for women is 26 years old, and average age for men is 32 years (Boletin, 1999).

The population pyramid for Argentina is weighted heavily towards the younger age groups. In 1999 Argentina's population aged 15-49 years represented 55% of the total population of the country. Gender is evenly distributed for this age category (Fact Sheet, 2000).

Table 7. HIV CASES BY YEAR OF TEST AND BY AGE CATEGORIES

	Age Categories						
Test Year	12-19	20-29	30-39	40-49	50+	Totals	% of Total
Unknowns						26	2%
1986	1					1	0.06%
1987		2				2	0.1%
1988	6	11	3	1		21	1%
1989	41	73	14	2		130	8%
1990	45	88	13	5	1	152	10%
1991	16	62	21	5	2	106	7%
1992	46	143	48	8	5	250	16%
1993	17	103	40	16	4	180	11%
1994	5	62	30	3	3	103	6%
1995	9	59	30	7	1	106	7%
1996	13	68	40	14	3	138	9%
1997	10	96	52	14	10	182	11%
1998	14	103	57	18	9	201	13%
Totals	223	870	348	93	38	1598	
% of Total	14%	54%	22%	6%	2%		

Table 7 shows the breakdown of HIV cases by age categories for the years 1986 - 1998. The youngest age category increased sharply in 1989, and then showed a consistent decrease from 1993 on. The 20-29 category fluctuated an increase in 1992 followed by a decrease in 1993 for several years, and then a return to gradual increases from 1996 on through the study period. The 30-39 age category has had similar fluctuations with gradual increases since 1996. Over 90% of the HIV+ cases are in the first three age groups, among women in their childbearing years.

Figure 2. STUDY CASES BY YEAR OF POSITIVE HIV TEST, WOMEN, BUENOS AIRES PROVINCE 1986 - 1998

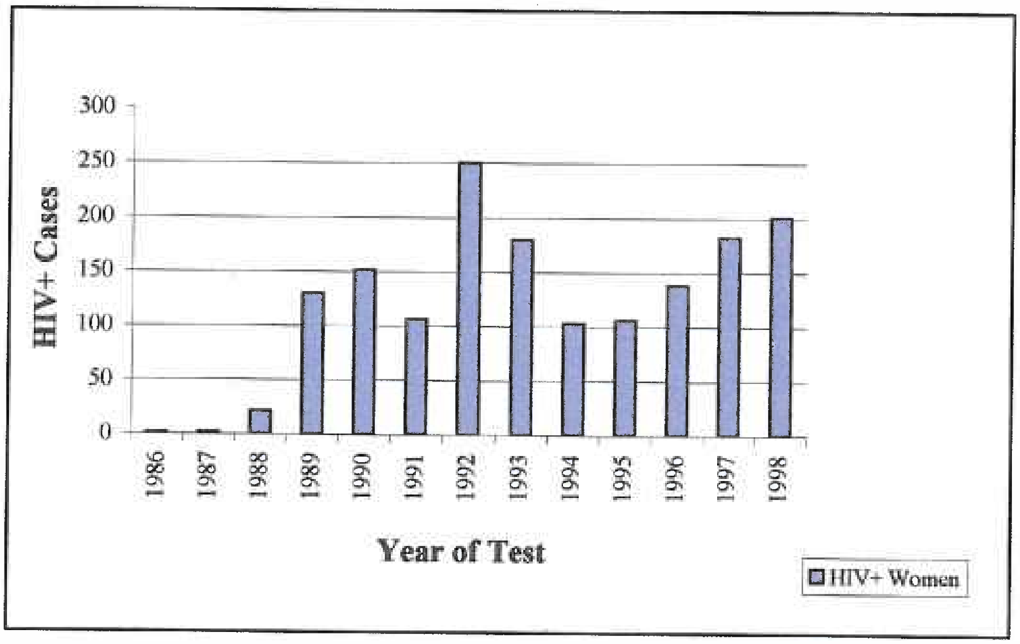


Figure 2 demonstrates the variable numbers of new HIV+ cases from the source data over the study period. Since 1994 the general direction has been increasing. The last year, 1998, finished with just 49 less than the highest year, 1992.

Cases have increased every year with the exception of 1991 and 1994. The sharp increase in 1992 is attributed to coinciding increased injection drug use and increased heterosexual transmission (PAHO, 1998).

Figure 3. CUMULATIVE STUDY CASES BY YEAR OF POSITIVE HIV TEST, WOMEN, BUENOS AIRES PROVINCE 1986 - 1998

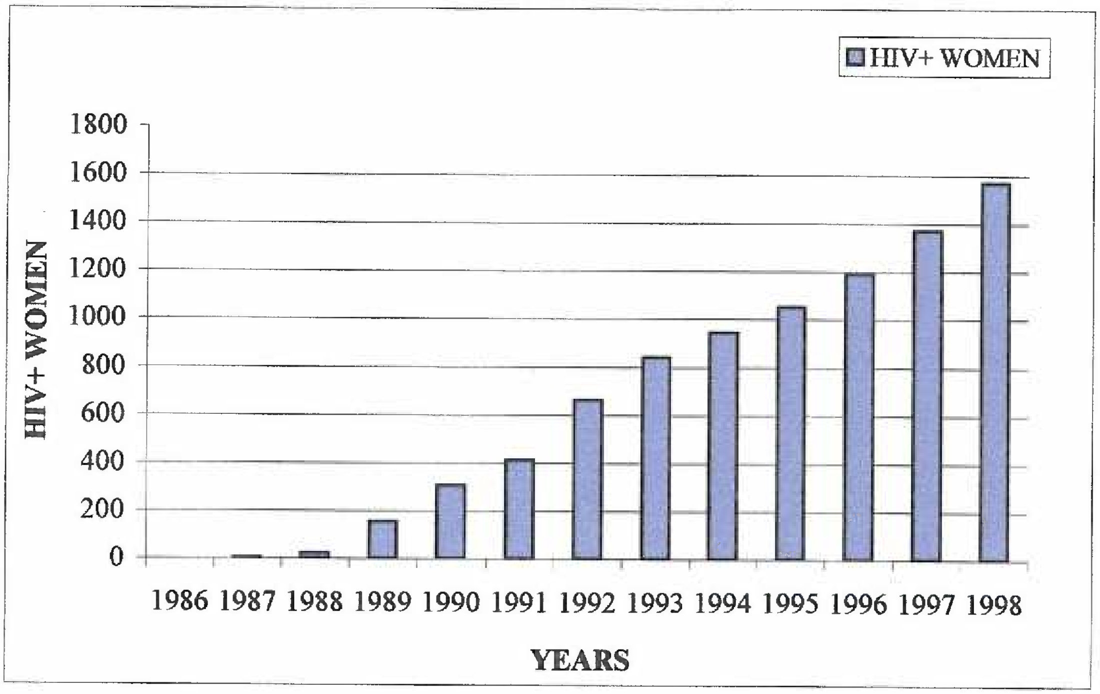
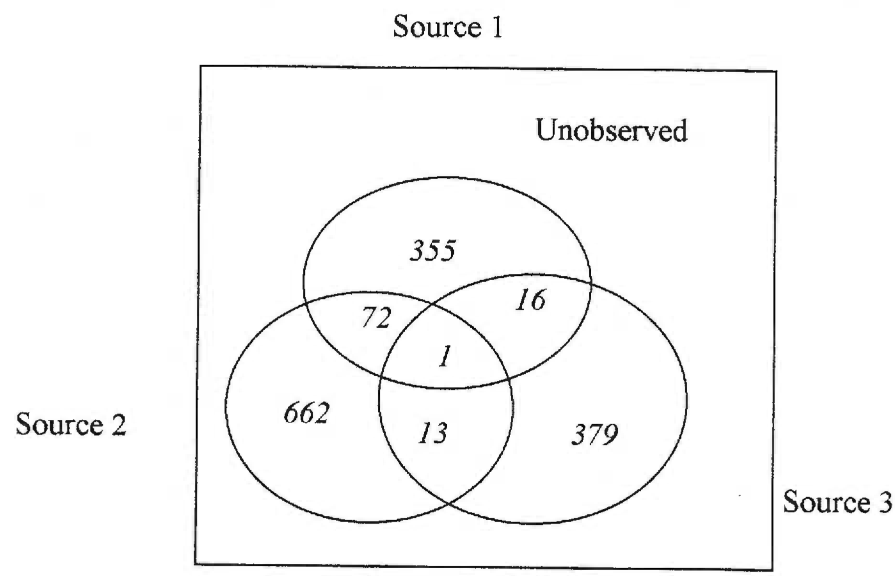


Figure 3 shows a generally steep climb in the number of cumulative cases. From 1989 to 1998 there is a twelve-fold increase of cases from 130 to 1598. Growth fluctuates from year- to - year as demonstrated in the tables.

Figure 4. ARGENTINE CAPTURE-RECAPTURE DATA DEMONSTRATING CROSSOVERS OF DATA



Source 1 (hospitals) = SJD + BGO + MP
Source 2 (laboratories) = BB + PHL
Source 3 (social service agencies) = IOMA + PAMI

Figure 4 presents the graphic representation of HIV cases by three categories of source data. There were 1498 observed women who were HIV+. Overlapping areas represent the recaptures between the different sources. Only one case was captured in all three sources. The area within the box, and outside of the circles represents the unobserved population. The best-fitting log linear model estimates the unobserved population to be an additional 8409 cases.

Table 8. INCOMPLETE MULTI-WAY CONTINGENCY TABLE WITH STUDY DATA

		Source 1			
		Yes		No	
		Source 2		Source 2	
		Yes	No	Yes	No
Source 3	Yes	a=1	b=16	e=13	f=379
	No	c=72	d=355	g=662	x=0

Table 8 demonstrates the design of the contingency table with "x" representing a structural zero rather than an actual zero. The structural zero is the unobserved population that has to be estimated through log linear modeling. This is different from traditional modeling techniques, where that cell would be treated as if it truly had zero observations.

Table 9. SPSS LOG LINEAR ANALYSIS TABLE

{value labels: yes=1; no=2}

Source 1	Source 2	Source 3	Count	Structure
1	1	1	1	1.00
1	2	2	355	1.00
2	1	1	13	1.00
2	2	1	379	1.00
2	1	2	662	1.00
1	1	2	72	1.00
1	2	1	16	1.00
2	2	2	0	0.00

Table 9 is the SPSS log linear table using the numbers or count from the contingency table. The value labels are 1=yes for presence in a source, and 2= no presence in a source. There are eight possible combinations for three sources of data.

Table 10. LOG LINEAR MODELS AND CONFIDENCE INTERVALS FOR POPULATION ESTIMATES

Model	df	AIC	BIC	$\hat{x}$	$\hat{N}$	BIC wt.
Independent	3	41.4128	43.92	6291	7789	0
1-2 interaction	2	1.4236	3.1	13758	15256	0.21
1-3 interaction	2	40.8249	42.5	5773	7271	0
2-3 interaction	2	9.0876	10.8	4204	5702	0
1-2,1-3 interaction	1	-0.1932	0.64	19301	20799	0.73
1-2, 2-3 interaction	1	-1.879	-1.04	8409	9907	1.68
1-3, 2-3 interaction	1	-1.7289	-0.89	3264	4762	1.56
1-2,1-3,2-3 interaction	0	0	0	8302	9800	

Wtd BIC 10139.96

Confidence Intervals	$\hat{x}_{l95}$	$\hat{x}_{u95}$	$\hat{N}_{l95}$	$\hat{N}_{u95}$
Independent	5070	7806	6568	9304
1-2 interaction	9442	20049	10940	21547
1-3 interaction	4561	7308	6059	8806
2-3 interaction	3307	5345	4807	6843
1-2,1-3 interaction	11047	33722	12545	35220
1-2,2-3 interaction	5045	14018	6543	15516
1-3, 2-3 interaction	2505	4253	4003	5751
1-2,1-3,2-3 interaction	1404	49086	2902	50584

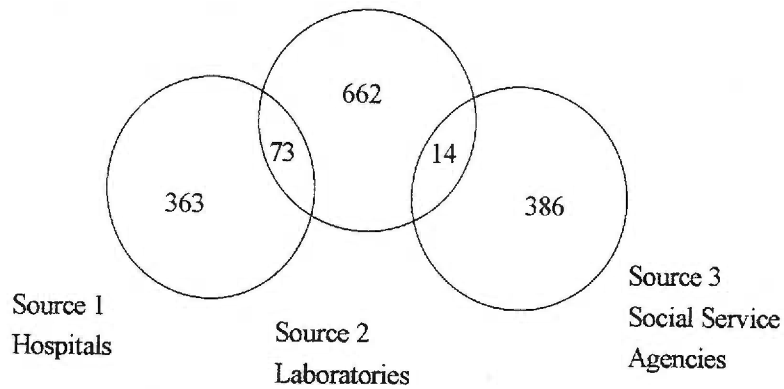
Table 10 shows the calculated values for the AIC, BIC, and estimates of the unobserved population. The confidence intervals are in the bottom half of the table. The 1-2, 2-3 interaction: hospitals(1)-laboratories(2), laboratories(2)-social service agencies(3) is the chosen model. The population estimate 95% confidence interval is

6,543 – 15,516 HIV+ women in the Buenos Aires Province. The weighted BIC average is 10,140 HIV+ women.

The chosen model with the 1-2, 2-3 interaction has a better fit as measured by the lowest values of both the AIC and the BIC. For the AIC and BIC to be a better fit than the model containing all three two-way interactions (saturated model), they have to be a value below zero. Since the saturated model has zero degrees of freedom, and the degrees of freedom are contained in the function of the AIC and BIC, they will also be zero.

The model with the second-best fit is the 1-3, 2-3 model. The AIC is –1.7289 and the BIC is –0.89 with a difference of only 0.15 from the best fitting model. However, there is considerable difference between their resulting estimates. The best-fit model estimate is 9907 and the second-best fit model estimate is 4762. The third best-fit model, the 1-2, 1-3 interaction, has an unobserved population estimate of 20,799. The AIC for the third choice is still in negative numbers below the saturated model, however the BIC is somewhat above zero at 0.64. The remaining models have increasing AIC and BIC with estimates ranging from 5702 to 15,256. All of the other models have estimates that fall between the best fitting model (9907) and the second-best fitting model (20,799). The weighted BIC average of 10,140 may be the best reflection of the true value of the unobserved population. It falls within the 95% confidence interval for the best-fitting model (6543-15516), and as an average, falls within the range of estimates of the other models.

Figure 5. MODEL DEMONSTRATING INTERACTIONS BETWEEN SOURCES



Log-linear model estimated number of missing cases = 8409
Population estimate = 9907
Model: 1 + 2 + 3 + 1-2 + 2-3
Deviance = .121 with 1 degree of freedom

Figure 5 illustrates the best-fitting model containing interactions between the laboratories and the hospitals (1-2), and between the laboratories and the social service agencies (2-3). The expected direction of the interaction or dependence would be a positive one. The public health laboratory (PHL) is a central referral laboratory for the province. It would be expected that HIV tests would be sent there for initial and/or confirmatory testing by hospitals and social service agencies.

DISCUSSION

Capture-recapture is a valuable tool when disparate data sets exist. Data from different sources with incomplete lists can be pooled. More of the available information, such as the duplicates, from the data sources is utilized to make an estimate of the size of the population. Capture-recapture does not require a complex or established reporting infrastructure for a disease entity. Depending on the population under scrutiny, sometimes unusual non-health related sources of data can be used, such as police records (opiate addicts, homeless, prostitution), homeless shelters (mentally ill), agencies (special groups that receive services), and attendance records (individuals belonging to treatment programs, children). Common health-related sources of data include hospitals, laboratories, pharmacies, doctors' offices and clinics, societies that offer support for a particular disease, and death records. In the current study sources used were: hospital records, laboratory results, and social service agencies. Only a unique identifier was necessary to follow individuals across sources, so confidentiality was maintained. .

Sub-populations within different socioeconomic and geographic groups may be heterogeneous, or have different chances of being identified by data sources. Larger sample sizes and pooling of data sources can help to minimize heterogeneity among sub-populations within the sample. In this study seven data sources were pooled into three categories.

Log-linear modeling helped to account for interactions and lack of independence between sources. The model with the best fit included two interactions: one between the hospitals and the laboratories, and one between the laboratories and the social service agencies. Using the best fitting model, which had both the lowest AIC and BIC value, the estimated prevalence of HIV is 9907 cases, (CI 95% 6,543 – 15,516). The BIC weighted estimate is 10,140 HIV+ women in the Buenos Aires Province. In the years 1989 – 1998 approximately 2264 AIDS cases were reported among women in the Buenos Aires Province (Boletin, 1999). Table 10 showed the next best fit according to the AIC and BIC values to be 4762 HIV+ women. In spite of the fit of this model, that estimate seems low, given that not all HIV or AIDS cases are captured. The UNAIDS Argentina Fact Sheet estimated 27,000 women in Argentina, ages 15 – 49 years, living with HIV/AIDS at the end of 1999. That translates to roughly 21,060 living in the Buenos Aires Province. The higher estimate from model number five (Table 10) of 20,799 cases is more consistent with UNAIDS estimate for HIV positive Argentine women.

There are a number of indicators that point to a maturing HIV/AIDS epidemic in Argentina. As demonstrated in Figures 2 and 3, the numbers of women who are infected are increasing sharply. Most evidence in mature epidemics demonstrates higher infection rates in young women than in young men. According to UNAIDS information, this is partly for physiological reasons, and partly because young women are more likely to have sex with older men, who may have had more exposure to infected partners in the past, (UNAIDS, 2000). Also, Argentine women are closing the gap as the proportion of male AIDS cases to women with AIDS has changed from 91:1 in 1987 to 2.6:1 in 1998.

In a mature epidemic HIV becomes established in the general population. High-risk sub-populations such as injection drug users and men who have sex with men may still contribute disproportionately to the spread of HIV. However, sexual networking in the general population becomes sufficient to sustain an epidemic independent of sub-populations at higher risk of infection. HIV prevalence consistently over one percent in the sentinel population of pregnant women can act as a numerical indicator, (UNAIDS, 2000). Among pregnant Argentine women, there have been varying numbers for prevalence. An earlier mentioned survey of pregnant Argentine women in 1998 found a median HIV prevalence of 0.29 and a maximum of 2.90% (Boletin, 1999). Another source quoted a range of 0.34% to 1.2% with the median at 0.93% (Fact Sheet, 2000). Table 6 shows that 76% of the HIV+ women identified in my study were in the childbearing years 20-39. It is unknown how many may have been pregnant. However, it is clear that they represent a large target group for interventions.

As the epidemic matures in Argentina, new infections will be concentrated in people who are close to the start of their sexual lives (UNAIDS, 2000). Table 6 shows the lowest age category of 12-19 years at 14% of the total, behind 20-29 at 54% and 30-39 at 22%. Table 7 shows the greatest numbers and growth of HIV+ cases among women in the 20-29 age group. There are important implications for the spread of HIV to children as women in this age group become infected with HIV. Testing and treatment of women and especially pregnant women could do much to reduce perinatal transmission.

Capture-recapture studies such as this one can and should be used to provide more accurate estimates of the population size of women who are HIV positive. The estimates can be used to target interventions for prevention and treatment of HIV in Argentine women. Since most cases of HIV in children occur at or around the time of birth or nursing, pregnant women should be tested routinely for HIV (Boletin, 1999). Testing of pregnant women is a World Health Organization recommendation that has implications for HIV in both women and children.

Capture-recapture can be used as a means of surveillance for HIV in pregnant women. It indicates that current AIDS data do not represent the number or trend of new cases of HIV among women. International organizations such as the World Health Organization (WHO) and the Joint United Nations Programme on HIV/AIDS (UNAIDS), are interested in global interventions for the control of HIV and AIDS. Surveillance of disease tailored to the epidemic in a country is recommended to concentrate data collection in populations most at risk of becoming newly infected with HIV. Surveillance can help to create a picture of the changes in the epidemic over time. Surveillance can give early warning of changes, can provide information for planning care and support, and can help determine the effectiveness of responses to the epidemic.

Future studies in Argentina would include a repeat study to identify trends over time among women. Risk factors and outcomes could be included and quantified. Socioeconomic and racial groups could be examined. A study to evaluate the impact and trends of HIV in children would be very interesting. Little data exists on children. Also a study to determine prevalence in other regions of Argentina which are not as populous or urban as Buenos Aires could help to describe the epidemic more completely.

Limitations

Capture-recapture studies, in general, are vulnerable to error through failure to meet the necessary assumptions. A closed population without individuals moving in or out of the studied area is not possible for most human populations. People can cease to be part of the population by means other than migration. In the case of HIV, a person may not be diagnosed until late in their disease process and die before they would have been seen at other facilities. Some of the lists that I used had noted that persons had died. In an urban setting, such as Buenos Aires and La Plata, immigration into the area to seek diagnosis and treatment would suggest the lack of a closed population. However, the Buenos Aires Province is the most urban and populous province in the country. The other population centers in the country are much smaller and long distances from each other and from Buenos Aires. Travel outside of the province to seek treatment seems less likely.

The second assumption of individuals being identifiable from one source to another is generally not a significant problem in human populations. Identification in this study included initials and the date of birth. Since I used handwritten logs, human transcription errors were possible. Also, people can change their names, and the letters of the alphabet are written a little differently in Argentina. However, when I was unsure, I would ask for help. I collected and entered all of the data myself which reduced the potential error due to handwriting interpretation.

The assumption that each individual has the same chance of appearing on any list can not be assured. Socioeconomic and geographic issues can effect the subpopulations within a sample. This study included only women, and was within one large, populous geographic area, the province of Buenos Aires. I don't think people with HIV would be likely to leave the province. Most people do not have their own cars, and fuel is very expensive, about four dollars a gallon. There are other costs to travel outside of the province with a car. Some of the major highways have a substantial toll. There is a good public transportation system, however, a sick person may not want to travel around.

Healthcare payment is a mix of private insurance and public. The facilities used as data sources were both public and private. Most of the cases came from public sources. Sources were pooled in the hopes of alleviating some of the heterogeneity problems.

The last assumption that capture on one list does not effect capture on another can not be ruled out. However, this potential limitation can be accommodated by modeling. With three sources of data, there are eight possible models. All eight were calculated and compared. Dependence was identified between the hospitals and the laboratories (1-2 interaction), and the laboratories and the social service agencies (2-3 interaction).

All three of the two-way interaction models had AIC values below the saturated model (less than 0). The three varied from a low estimated total (the 1-3, 2-3 interaction) of 4762 (CI 95% 4003 – 5751), to a high estimate (the 1-2,1-3 interaction) of 20,799 (CI 95% 12,545 – 35,220). The chosen model is positioned between these two. The chosen model is a 1-2,2-3 two-way interaction. The estimate was a total of 9907 (CI 95% 6543 – 15,516). The weighted BIC estimate is 10,140.

SUMMARY AND CONCLUSIONS

This study of HIV prevalence among Argentine women in the Buenos Aires Province from 1989 to 1998 produced an estimate of 9907 cases of HIV positive women (CI 95% 6543 –15,516). The count of AIDS cases among women in the province is about 2269 for that time period (Boletin, 1999). The difference is an additional 7638 women who are HIV positive, but have not yet been counted among the AIDS cases. Several of the models had higher estimates than the best-fitting model. The BIC weighted average estimate would add another 233 cases to the 7638 figure. The data shows that there are significantly more women who will develop AIDS in the next decade, and that the numbers of HIV positive women are increasing for the time period.

Capture-recapture is a practical method to estimate the prevalence of HIV in this setting. Since capture-recapture relies on existing data sources, it is relatively inexpensive to do. A study can be undertaken relatively quickly to produce timely data. Capture-recapture can provide accuracy using weighted estimates and confidence intervals. It has been used in many countries around the world. It is especially useful for countries that do not have a strong public health infrastructure. Capture-recapture has also been used successfully to estimate elusive populations.

In this study high quality data on HIV positive women existed at a number of sites. With more time, I think more sources could have been identified. Using existing data to estimate cases of HIV among women gives a more accurate and timely picture of the direction of the HIV epidemic and patterns of transmission. This study indicates that HIV is increasing in Argentine women. The greatest numbers of HIV cases were found in women in their reproductive years. This means that HIV will also increasingly infect children at the time of birth and during nursing. In accordance with WHO recommendations, pregnant women should be offered voluntary testing for HIV regardless of risk factors. Those women with HIV infection, if tested and treated, can live healthier longer lives, make better decisions about pregnancy, and greatly reduce the transmission of HIV to children. Capture-recapture can be used to obtain more accurate estimates of population size for women with HIV, and can lead to improvements in planning for prevention and treatment of HIV for women and children.

REFERENCES

Ackman DM, Birkhead G, Flynn M. Assessment of surveillance for meningococcal disease in New York State, 1991. Am J Epi 144(1): 78-82, 1996.

Ballivet S, Salmi LR, Dubourdieu D. Capture-recapture method to determine the best design of a surveillance system. Application to a thyroid cancer registry. Eur J Epidemiol Feb 16(2): 147-153, 2000.

Bernhardson W. Argentina, Uruguay, and Paraguay, a lonely planet travel survival kit. Lonely Planet Publications, 1996.

Betts CD, Astarloa L, Bloch C, Zacarias F, Chelala C. The changing face of AIDS in Argentina. JAMA 276(2): 94-96, 1996.

Bishop Y, Fienberg SE, Holland PW. Discrete multivariate analysis: theory and practice. Cambridge, MA, MIT Press, 229-256, 1975.

Boletin sobre el sida en la Argentina. Ministerio de Salud y Accion Social de la Nacion. Ano VI, Numero 16, Marzo 1999.

Boyer PJ, Dillon M, Navaie M, Deveikis A, Keller M, O'Rourke S, Bryson YJ. Factors predictive of maternal-fetal transmission of HIV-1, Preliminary analysis of zidovudine given during pregnancy and/or delivery. JAMA 271(24): 1925-1930, 1994.

Bruno G, LaPorte RE, Merletti f, Biggeri A, McCarty D, Pagano G. National diabetes programs, application of capture-recapture to count diabetes? Diabetes Care 17(6): 548-56, 1994.

Cates, Jr. W, Smith JC, Rochat RW, Patterson JE, Dolman A. Assessment of surveillance and vital statistics data for monitoring abortion mortality, United States, 1972-1975. Am J Epi 108(3): 200-06, 1978.

Chang YF, LaPorte RE, Aaron DJ, Songer TJ. The importance of source selection and pilot study in the capture-recapture application. J Clin Epi 52(10): 927-928, 1999.

Chang YF. Capture-recapture history. www.pitt.edu/~yuc2/cr/history.htm. March,1999.

Chiu WT, Dearwater SR, McCarty DJ, Songer TJ, LaPorte RE. Establishment of accurate incidence rates for head and spinal cord injuries in developing and developed countries: a capture-recapture approach. J of Trauma 35(2): 206-211, 1993.

Cormack RM. Log-linear models for capture-recapture. Biometrics 45: 395-414, 1989.

Cormack RM. Problems with using capture-recapture in epidemiology: an example of a measles epidemic. J Clin Epi 52(10): 909-914, 1999.

Cormack RM. Reply to preceding comments. J Clin Epi 52(10): 929-933, 1999.

Domingo-Salvany A, Hartnoll RL, Maguire A, Suelves JM, Anto JM. Use of capture-recapture to estimate the prevalence of opiate addiction in Barcelona, Spain, 1989. Am J Epi 141(6): 567-574, 1995.

Domingo-Salvany A, Hartnoll RL, Maguire A, Brugal MT, Albertin P, Cayla JA, Casabona J, Suelves JM. Analytical considerations in the use of capture-recapture to estimate prevalence: case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. Am J Epi 148(8): 732-739, 1998.

Douglas GC, King BF. Maternal-fetal transmission of human immunodeficiency virus: a review of possible routes and cellular mechanisms of infection. CID 15(October): 678-691, 1992.

Draper D. Assessment and propagation of model uncertainty. J R Stat Soc [B]57:45-70, 1995.

Editorial. HIV in Latin America and the Caribbean. Lancet 355(9221): 2087, June 17, 2000.

Egeland GM, Perham-Hester KA, Hook EB. Use of capture-recapture analyses in fetal alcohol syndrome surveillance in Alaska. Am J Epi 141(4): 335-341, 1995.

Fienberg SE. The multiple recapture census for closed populations and incomplete 2k contingency tables. Biometrika 59: 591-603, 1972.

Fisher N, Turner SW, Pugh R, Taylor C. Estimating numbers of homeless and homeless mentally ill people in northeast Westminster by using capture-recapture analysis. BMJ (308): 27-30, 1994.

Frischer M, Bloor M, Finlay A, et al. A new method of estimating prevalence of injecting drug use in an urban population, results from a Scottish city. Int J Epi 20: 997-1000, 1991.

Green TA, Karon JM, Nwanyanwu OC. Changes in AIDS incidence trends in the United States. J Acquired Immune Deficiency Syndromes 5(6): 547-555, 1992.

Havlir DV, Richman DD. Viral dynamics of HIV: implications for drug development and therapeutic strategies. Ann Intern Med 124: 984-994, 1996.

Hewitt D, Milner J. Prevalence of hemophilia in Ontario, 1966. Can Med Assoc J 102: 174-7, 1970.

Hook EB, Regal RR. The value of capture-recapture methods even for apparent exhaustive surveys. The need for adjustment for source of ascertainment intersection in attempted complete prevalence studies. Am J Epi 135(8): 1060-1067, 1992.

Hook EB, Regal RR. Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. Am J Epi 137(10): 1148-1166, 1993.

Hook EB, Regal RR. Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. Am J Epi 142(9, Suppl.): S48-S52, 1995.

Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epi Rev 17(2): 243-264, 1995.

Hook EB, Regal RR. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. J Clin Epi 52(10): 917-926, 1999.

International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: history and theoretical development. Am J Epi 142(10): 1047-58, 1995.

International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: applications in human diseases. Am J Epi 142(10): 1059-68, 1995.

Ismail AA, Beeching NJ, Gill GV, Bellis MA. How many data sources are needed to determine diabetes prevalence by capture-recapture? Int J Epi Jun 29(3): 536-41, 2000.

Kozinetz CA, Skender ML, MacNaughton N, et al. Epidemiology of Rett syndrome: a population-based registry. Pediatrics 91: 445-50, 1993.

Lallemant M, LeCoeur S, Tarantola D, Mann J, Essex M. Antiretroviral prevention of HIV perinatal transmission. Lancet 343(8910): 1429-1430, June 4, 1994.

LaPlace PS. Sur les naissances, les mariages, et les morts. (In French). In: Historie de l'Academie Royale des Sciences, Annee 1783, Paris: p. 693 (cited by Seber GAF, The estimation of animal abundance and related parameters, 1982.)

LaPorte RE, McCarty DJ, Tull ES, Tajima N. Counting birds, bees, and NCDs. The Lancet 339: 494-5, 1992.

LaPorte RE, Tull ES, McCarty D. Monitoring the incidence of myocardial infarctions: applications of capture-mark-recapture technology. Int J of Epi 21(2): 258-62, 1992.

Mastro TD, Kitayaporn D, Weniger BG, Vanichseni S, Laosunthorn V, Uneklabh T, Uneklabh C, Choopanya K, Limpakarnjanarat K. Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. Am J Pub Hlth 84(7): 1094-1099, 1994.

McCarty D, Kwoh CK, LaPorte RE. The importance of incidence registries for connective tissue diseases (an editorial). J of Rheum 19(1): 1-7, 1992.

McCarty D, Tull ES, Moy CS, Kwoh CK, LaPorte RE. Ascertainment corrected rates: applications of capture-recapture methods. Int J of Epi 22(3): 559-65, 1993.

McCarty D, Manzi S, Medsger, Jr. TA, Ramsey-Goldman R, LaPorte RE, Kwoh CK. Incidence of systemic lupus erythematosus, race and gender differences. Arth & Rheum 38(9): 1260-1270, 1995.

McGilchrist CA. Model selection and population size using capture-recapture methods. J Clin Epi 52(10): 915, 1999.

Modesitt SK, Hulman S, Fleming D. Evaluation of active versus passive AIDS surveillance in Oregon. Am J Pub Hlth 80(4): 463-464, 1990.

Palfrey JS, Walker DK, Haynie M, et al. Technology's children: report of a statewide census of children dependent on medical supports. Pediatrics 87:611-18, 1991.

Pan American Health Organization/World Health Organization (PAHO). AIDS surveillance in the Americas. May 2000.

Pollock KH. Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations; past, present, and future. J Am Stat Assoc 86: 225-38, 1991.

Prevots DR, Sutter RW, Strebel PM, et al. Completeness of reporting for paralytic poliomyelitis, United States, 1980 through 1990: implications for estimating the risk of vaccine-associated disease. Arch Pediatr Adolesc Med 148:479-85, 1994.

Raggi R, Blanco GA. Epidemic model of HIV infection and AIDS in Argentina. Status in 1990 and predictive estimates. Medicina (B.Aires) 52(3): 225-35, 1992.

Sanghavi DM, Gilman RH, Lescano-Guevara AG, Checkley W, Cabrera LZ, Cardenas V. Hyperendemic pulmonary tuberculosis in a Peruvian shantytown. Am J Epi 148(4): 384-389, 1998.

Seber GF. The estimation of animal abundance and related parameters, 2nd edn. London: Griffin & Co. 1982.

Sekar DD, Deming WE. On a method for estimating birth and death rates and the extent of registration. Am Stat Assoc J 100-15, 1949.

Snell J. The looming threat of AIDS and HIV in Latin America. Lancet 354(9185): 1187, October 2, 1999.

Tarantola DJM, Mann JM. Global Expansion of HIV infection and AIDS. Hosp Prac October 15, 1996, 63-79.

UNAIDS/WHO AIDS epidemic update: December 2000.

UNAIDS/WHO Epidemiological fact sheet on HIV/AIDS and sexually transmitted infections, Argentina. 2000.

UNAIDS/WHO Guidelines for Second Generation HIV Surveillance. Second generation surveillance for HIV: the next decade. 2000.

UNAIDS/WHO Statement: Prevention of HIV transmission form mother-to-child: meeting on planning for programme implementation. Geneva, March 23-24, 1998.

Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. J Chron Dis 27: 25-36, 1974.