V. Projection of AIDS Incidence in Women in New York State

Lawrence Lessner, PhD, MPH

Introduction

One objective of the New York State Newborn HIV Seroprevalence Study was to provide information for AIDS case projections. Earlier projections were developed by linear extrapolation from reported AIDS cases. The limitations of using AIDS case reports include underreporting of AIDS cases, reporting delays, and, most importantly, the long incubation period of HIV infection.

To improve the projection process, we turned to data from two other sources: fertility data, and the results of the Newborn HIV Seroprevalence Study. With these and other data, we estimated the number of HIV-infected women of reproductive age in 1988 and then the number of women infected each year during 1981–88. From this distribution of the incidence of infection and an incubation distribution, we projected the anticipated AIDS case incidence for this population.

The AIDS case projection methods developed here are a form of back calculation.[1,2] The major advantage of this application is that information from another data source, newborn HIV seroprevalence, is employed to obtain an estimate of the number of infected women of childbearing age in 1988.

As newborn serosurveys for antibody to HIV are implemented in more states, this type of analysis will become increasingly important in efforts to anticipate future AIDS cases.

Methods

AIDS case projections were developed in three steps:

1. Newborn seroprevalence results from 1988 provided an estimate of the number of infected women of reproductive age in 1988.

2. An incubation distribution was used to create an expected case series for each number of a family of hypothetical distributions of incident infections. A "best distribution" of infections was selected by comparing the expected and actual case series.

3. A lower bound on future AIDS cases for this population was obtained from the estimates of presently infected women.

Estimating the Number of Infected Women

The number of HIV-seropositive newborns in 1988 was obtained from the Newborn HIV Seroprevalence Study[3] for each of 80 age, racial/ethnic, and geographic strata. The number of HIV-infected women of reproductive age (15–44 years) in each stratum was determined by dividing the number of newborn HIV seropositives[3] by the fertility rate for that stratum. The total estimate is the sum of these 80 stratum-specific estimates.

The fertility rate R of a given female population is defined as B, the number of live births that occur in the population during the given year, divided by P, the average or midyear population for the given female subgroup of age 15 to 44 years: R = B/P. An estimate of P_i, the size of the infected female population, was obtained by inverting the relationship: $R_i = B_i/P_i$. If B_i is the number of seropositive births and R_i is the fertility rate in the infected female population, the population size is $P_i = B_i/R_i$. Assuming the same fertility rate for infected women as for the entire stratum $R_i = R$, then $P_i = B_i/R$.

Since the fertility rate R varies by age, race, and geographic location, the population was stratified by age (15–19, 20–29, 30–39, 40–44 years), race (Hispanic, non-Hispanic Black, non-Hispanic White, and all others), and five geographic regions: New York City (the five bor-

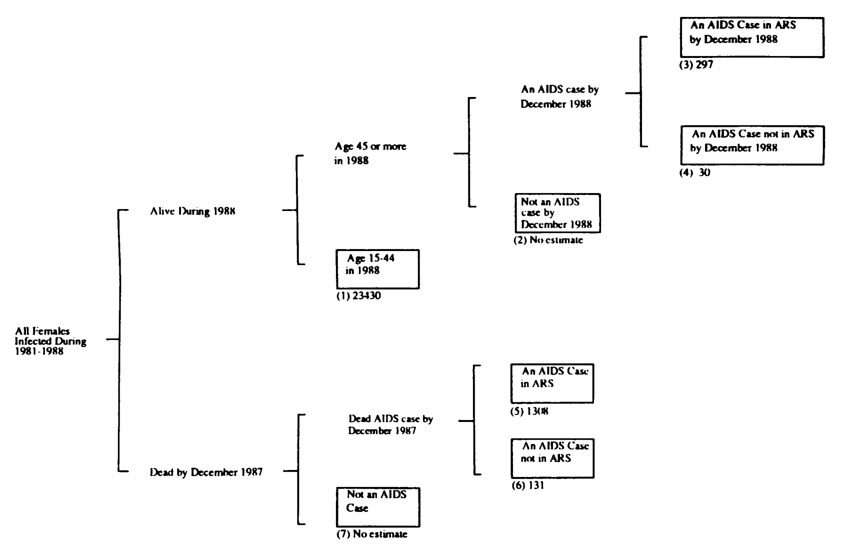

FIGURE 1—Procedure for Estimating the Total Number of Women Infected During 1981–88

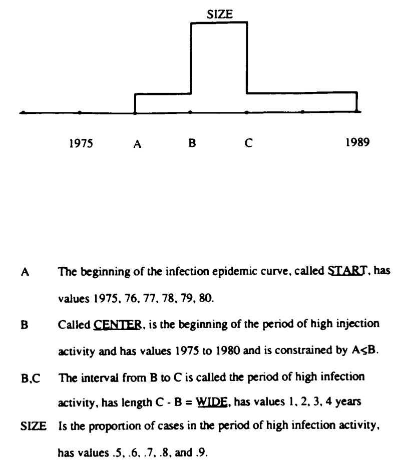

FIGURE 2—Generic Candidate For Distribution of Infections

oughs); New York City Vicinity (Nassau, Suffolk, and Westchester Counties), Mid-Hudson Valley (Dutchess, Orange, Putnam, Rockland, Sullivan, and Ulster Counties), Upstate Urban (Albany, Schenectady, Onondaga, Erie, and Monroe Counties), and Upstate Rural (all other counties).

Vital statistics data for 1980 and 1988 were used to obtain estimated stratum-specific fertility rates. The 1980 fertility rates were increased in each stratum to reflect the 1988 fertility experience: ages 15–19 were increased by 17.8 percent, ages 20–29 by 4.1 percent, ages 30–39 by 30.5, and ages 40–44 by 48.9 percent (J. Cooney, Bureau of Production Systems Management, New York State Department of Health, oral communication, January 1990).

The model for estimating the number of women infected during 1981–88 is presented in Figure 1. This tree diagram partitions the ever-infected population into disjoint subgroups for which infection is estimated. Subgroups 4 and 6 use estimates from the AIDS Reporting System (ARS) which is considered 80 to 90 percent complete (Dr. Meade Morgan, Chief of Statistical and Data Management Branch, Division of HIV/AIDS, Centers for Disease Control, oral communication, January 19, 1990).

Estimating the Distribution of Infections

A family of candidate distributions was examined for their utility in describing the incidence of HIV infection in women during 1981–88. The generic candidate (Figure 2), a two-step histogram, was generated by specifying four parameters: START, WIDE, CENTER, and SIZE. START, the beginning of the infection epidemic, had possible values 1975, 1976, . . ., 1980. The interval of high infection activity begins at CENTER and had possible values that range from START to 1988. WIDE is the width of the period of high infection activity and had possible values 1, 2, 3, or 4 years. SIZE is the proportion of infections within the period of intense infection activity and had possible values 0.5, 0.6, 0.7, 0.8 or 0.9. With this model there are 1,200 possible two-step histograms.

A Weibull incubation distribution was selected with a median of 10 years and a fifth percentile of 3.5 years. This selection was based on reports of incubation distributions in a hemophiliac population and in the San Francisco Gay Cohort,[4,5] with estimated median incubation times of 10.0 and 9.8 years, respectively. The estimated fifth percentiles of incubation times were 3.6 and 3.4 years, respectively. An expected case series was created for each of the 1,200 candidates and compared to the annual observed, lag-adjusted AIDS case incidence.

The comparison measure, called ERROR, is the sum of the absolute differences between the observed and expected incidences for each year in each candidate series. The 1,200 ERROR sums were ordered from smallest to largest. Absolute accuracy was measured by ERROR. Relative accuracy was obtained with %ERROR, which is ERROR divided by 2,589, the total incidence observed during 1981–88.

To validate the use of an incubation distribution with a median of 10 years and a fifth percentile of 3.5 years, a family of nine Weibull incubation distributions was created having a 50th percentile of 9, 10, or 11 years and a 5th percentile of 3.0, 3.5 or 4.0 years. Each of these incubation distributions and each of the 1,200 histograms of infection were combined to produce 10,800 candidates, and for each candidate an expected case series was calculated. The top 50 candidates, as ranked by ERROR, were examined to find the smallest ERROR for each incubation distribution (Table V-1). This calculation was at least consistent with the choice of 10 and 3.5 years, although the choices (10, 4.0) and (9, 4.0) appeared equally good. The choices where the 50th percentile is 11 or the 5th percentile is 3.0 appeared considerably less consistent with the observed case series.

Projections of AIDS Case Incidence in Women

The candidate with the minimum value of ERROR was selected as the estimate of the distribution of infections. The expected number of cases was calculated for each year as a lower bound on the anticipated future incidence of AIDS cases in women.

TABLE V-1—Comparison of Incubation Distribution Parameters and ERROR of the Best Candidate

Candidates for the 50th Percentile	Candidates for 5th Percentile 3.0	3.5	4.0
9	(*)	70.9	58.9
10	123.7	61.0	61.9
11	121.4	91.2	105.8

(*) at least 154.5

NOTE: Entries are the ERROR for the best candidate for each combination of parameters

TABLE V-2—Summing the Decision Tree (Figure 1)

Box	Population (Origin of the estimate)	The Estimate
1	Women of childbearing age in 1988 (Inverting the fertility rates over 80 strata)	23,430
2	Alive in 1988 and age 45 or more, but not an AIDS case by Dec. 1988 (Unknown)	No estimate
3	An AIDS case in ARS by Dec. 1988 (ARS)	297
4	An AIDS case by Dec. 1988 but not in ARS (10% of 297)	30
5	AIDS case in ARS and dead by 1987 (ARS)	1,308
6	Dead AIDS case by Dec. 1987 and *not* in ARS (10% of 1,308)	131
7	Not AIDS case and died from either other causes or HIV-related disease not AIDS (Unknown)	No estimate
Total		25,196

Results

Estimates of Number of Infected Women (15–44 years)

If we assume equal fertility rates for the general and infected populations in each of 80 age, racial/ethnic, and geographic strata, an estimated total of 23,430 women age 15–44 years were infected with HIV in New York State in 1988. This total was used to obtain an estimate of all women infected during 1981–88 (Figure 1, Table V-2).

Estimates were not possible for certain populations. Box 2 represents all women infected during this period who were alive and age 45 years or more in 1988 but who were not a defined Centers for Disease Control (CDC) case by December 1988. This is the only source of infected women who are potential future AIDS cases for which there is no estimate. Box 7 represents all women infected during this period who died by December 1987 from causes not related to HIV infection. This population will not influence future AIDS incidence.

The estimate of all HIV-infected women of reproductive age in New York State during 1981–88 is 25,196.

Estimates of Distribution of Infections during 1981–88

The 1,200 candidates were ordered by increasing values of ERROR, and the observed and expected cases for six representative candidates were compared (Table V-3). The best candidate had half the ERROR of the next-best candidate.

Three populations of candidates for the infection distribution—the best 10, 20, and 30 candidates—were examined for consensus on the four parameters (Table V-4). START had a clear single mode across all three populations. This is a strong consensus that year 1980 was the beginning of the infection distribution. For CENTER there was a strong consensus across all three populations for 1984 and 1985.

The consensus for the percentiles (Table V-5) shows that the cumulative proportion of new infections nearly doubles in 1984 for each of the three populations. Thus the period of intense infection activity began in 1984. The three-year period 1984–86 accounts for 70 to 80 percent of all infections.

Projections of Future AIDS Cases

The three best candidates for the distribution of infections were used with the incubation distribution to project future AIDS case incidence for 1989 to 1993 (Table V-6). A considerable consensus was found for each year across the three best candidates, and the means and standard deviations for the top 20 candidates substantiated this consensus. The average observed coefficient of variation for these five years was 0.0313. The best candidate indicates that if there were no additional infections after December 1988, the minimum cumulative number of cases expected December 1993 is 8,295.

The linear projections in Table 6 are estimates of future AIDS incidence calculated from actual monthly AIDS surveillance data by linear extrapolation.[6] These linear projections and the best candidate show considerable similarity for the first three years. Beginning in 1992 there is a growing divergence. The linear projections continue to increase each year by a fixed amount, 277 cases. The rate of increase in the candidates is governed by the distribution of infections and, in time, if one assumes no additional infections, there will be a slowing in the rate. Since these projections make no attempt to anticipate future infections, they must be viewed as a conservative estimate.

Discussion

These projections of future AIDS cases among women of reproductive age are based on the results of a newborn HIV seroprevalence study in New York State. An earlier method, linear extrapolation, yielded similar projections for the short term.[6] Projections by extrapolation will always exceed those produced with a method that uses a fixed number of infected because the projections of AIDS cases produced here are based only upon presently infected persons; they must be considered a lower bound of expected cases.

Back calculation is a general class of methods that uses an incubation distribution and the observed case incidence series to estimate the distribution of infections over time.[1,2] The method discussed here, employing two-step histogram, suffers from all the usual disadvantages of back-calculation methods: underreporting of case incidence, limited knowledge of the incubation distribution, and absence of case reports before a certain year (in this case 1981). The key theoretical advantage of back calculation, compared to earlier empiric extrapolation, is its direct quantification of the size of a previously infected population.

The potential limitations of the validity of the work presented here include the assumption of equal fertility between the general and the infected populations in each stratum, the selection of an incubation distribution, the simplicity of the two-

TABLE V-3—Expected Case Series for Selected Best Candidates, Assuming 25,196 Total Infected Woman and a Weibull Incubation Distribution Having 50th Percentile of 10 Years and 5th Percentile of 3.5 Years

Year	Observed	Best	Best 2nd	Best 3rd	Best 10th	Best 20th	Best 30th
1981	9	18	29	32	19	23	21
1982	35	50	58	65	52	62	58
1983	96	99	98	110	104	124	116
1984	155	168	159	168	183	210	196
1985	267	269	262	253	305	319	298
1986	435	436	420	411	476	458	426
1987	658	670	637	644	697	649	595
1988	934	939	901	915	957	889	810
Total	2589	2649	2564	2598	2793	2734	2520
ERROR		61.04	121.8	152.07	204.41	253.97	323.23
%ERROR		2.40	4.7	5.90	7.90	9.80	12.50

TABLE V-4—Consensus for Distributional Parameters in the Best 10, 20, and 30 Candidates

Variables	Best 10	Best 20	Best 30
START			
75	1	1	1
76	0	1	1
77	0	2	2
78	0	2	5
79	3	4	7
80	6	10	14
SIZE			
0.5	1	2	2
0.6	3	3	5
0.7	2	5	10
0.8	3	7	9
0.9	1	3	4
WIDE			
1	2	6	8
2	2	3	5
3	4	7	12
4	2	4	5
CENTER			
84	5	10	12
85	4	8	14
86	1	2	4

step histograms, and the underrepresentation of infections occurring during the survey period, 1988.

A fundamental assumption in relating newborn seropositivity to infections in all women of reproductive age is that HIV-infected women have the same fertility rate as uninfected women in the same age, racial/ethnic, and geographic strata. However, the possible effects of HIV infection challenge this assumption. Intravenous drug use could also reduce fertility and thus low-bias the estimate of infected women of childbearing years.

Few published reports address the potential differences in fertility between the general population and HIV-infected or intravenous-drug-using women. Commonly the comparison is between seropositive and seronegative women who are also or have been intravenous drug users.[7–10] For these groups no differences in fertility have been observed.

If there is a decrease in fertility in HIV-infected women, it may be secondary to associated intravenous drug use. Selwyn[7] compared the pregnancy rate of women in the Bronx in 1985 with a cohort of intravenous drug users enrolled in a methadone maintenance treatment program. The latter group had an estimated decrease in the fertility rate of 4.7 pregnancies per 100 person-years. Thus if there is a decrease in fertility, it is likely to be small (Dr. Peter Selwyn, Department of Epidemiology and Social Medicine, Montefiore Medical Center, New York, NY, oral communication, January 1990).

To investigate the sensitivity of the assumption of equal fertility, we computed the estimate of total infected women for progressively decreasing fertility rates in the infected population. If

$$\sum_{i=1}^{80} \frac{B_i}{R_i}$$

is the original estimate of the total number of infected women of childbearing age in 1988,

$$\sum_{i=1}^{80} \left[\frac{B_i}{(1-a)R_i} \right] = \left[\frac{1}{1-a} \right] \sum_{i=1}^{80} \frac{B_i}{R_i}$$

is the estimate resulting from a 100a percent decrease in fertility. The results of these calculations are given in Table V-7, as well as the ERROR for the best, the best five, and the best 10 candidates for each scenario of decreased fertility. These ERROR values measure the consistency of the assumed decrease in fertility.

As fertility decreases, the estimated total number of infected women increases to 48,626 (corresponding to a 50 percent decrease in fertility). The three ERROR values generally increase with decreases in fertility. This generally increasing ERROR implies a degradation in the consistency of the model and casts doubt on the estimated total number of infected women that results from a severe decrease in fertility. Modest decreases in fertility result in larger but still credible estimates of infecteds (a 15 percent decrease in fertility results in a 16 percent increase in infecteds and a 13 percent increase in best-10 ERROR). Larger decreases in fertility are clearly much less credible than the estimate of no difference in fertility rates between the HIV-infecteds and the general population.

The selection of the incubation distribution was based upon two articles[4,5] that indicated consensus in the choice of a Weibull distribution and its parameters.

TABLE V-5—Consensus for Percentiles of Incident Infections in the Best 10, 20, and 30 Candidates

Year	Best 10	Best 20	Best 30
1980	.06 (.01)	.06 (.01)	.06 (.02)
1981	.10 (.02)	.10 (.03)	.10 (.03)
1982	.15 (.04)	.13 (.04)	.14 (.04)
1983	.19 (.06)	.17 (.06)	.18 (.06)
1984	.34 (.07)	.32 (.09)	.30 (.09)
1985	.61 (.12)	.67 (.16)	.63 (.17)
1986	.87 (.09)	.88 (.10)	.86 (.11)

NOTE: The average cumulative number of new infections from 1975 to the given year for the Best 10, 20 and 30 Candidates is shown. The standard deviation for the data is given in parentheses

Attempts at verification were consistent with those reported in both articles.

Although structurally more complex histograms would yield more accurate results, our simple histograms still offer a wide variety of choices. In any case, parametric modelers performing a back calculation with maximum likelihood methods must necessarily limit the form of their candidate populations for each estimate.[2] The accuracy of the expected case series in the best several candidates (Table V-3), the strength of consensus in parameter estimates (Table V-4), and the strength of consensus in the future num-

TABLE V-6—Projections of Future AIDS Cases from Seroprevalence Data and by Linear Extrapolation

Year	Linear Projections*	Seroprevalence Projections Best	2nd Best	3rd Best	Summary of the Top 20 Candidates Mean	SD
1989	1170	1221	1180	1199	1232	47.5
1990	1446	1487	1447	1469	1502	53.0
1991	1723	1714	1679	1702	1732	54.9
1992	1999	1885	1858	1880	1904	52.7
1993	2276	1988	1970	1989	2006	46.6
Total 1989–93	8614	8295	8134	8239	8376	

*Estimates of future AIDS incidence from an empiric model using actual monthly AIDS surveillance data.

TABLE V-7—Sensitivity and Consistency of the Estimate of the Total Number of Women Infected during 1981 to 1988 to Decreases in the Fertility Rate of Infected Women

% Decrease in Fertility of Infected Women	# Women of Childbearing Years	Total Number Infected during 1981–1988	ERROR Best	Best 5	Best 10
0	23430	25196	61	134	164
10	26033	27799	107	144	190
12	26565	28331	71	140	180
15	27565	29331	94	151	185
17	28118	29884	125	162	195
20	29288	31054	175	209	229
22	29897	31663	176	222	237
30	33472	35238	88	175	242
42	40217	41983	206	246	289
50	46860	48626	254	328	363

bers (Table V-6) confirm the adequacy of the simple histograms.

This method, utilizing HIV seroprevalence data from newborns, has resulted in improved projections of AIDS case incidence in women and may be useful in developing improved projections of pediatric AIDS.

Summary

The results of the Newborn HIV Seroprevalence Study in New York State for 1988 were used to estimate the number of human immunodeficiency virus (HIV)-infected women of childbearing age in 1988. The estimate was accomplished for each of 80 age, racial/ethnic, and geographic strata by dividing the number of seropositive newborns by the 1988 estimated fertility rate. Summing across strata yielded a total of 23,430 infected women age 15 to 44 years in 1988. From this estimate a tree model was used to estimate the number of women infected during 1981–88. Back calculation was used to distribute these infections over time, and the selected incubation distribution was used to estimate future AIDS cases for this population: 8,295 cases by 1993. □

References

1. Brookmeyer R, Gail MH: Methods for projecting the course of the AIDS epidemic. JNCI 1988; 80:900–911.
2. Brookmeyer R, Damiano A: Statistical methods for short-term projections of AIDS incidence. Stat Med 1989; 8:23–34.
3. Novick LF, Glebatis DM, Stricof RL, MacCubbin PA, Lessner L, Berns D: Newborn seroprevalence study: Methods and results. Am J Public Health 1991; 81(suppl): 15–21, Ch. II.
4. Brookmeyer R, Goedest JJ: Censoring in an epidemic with an application to hemophilia-associated AIDS. Biometrics 1989; 45:325–335.
5. Bacchetti P, Moss AR: Incubation period of AIDS in San Francisco. Nature 1989; 338:251–253.
6. New York State Department of Health: AIDS in New York State. Albany, NY: New York State Department of Health, 1988.
7. Selwyn PA, Schoenbaum EE, Davenny K, *et al:* Prospective study of human immunodeficiency virus infection and pregnancy outcome in intravenous drug users. JAMA 1989; 261:9:1289–1294.
8. Johnstone FD, MacCallum L, Brettle R, Inglais JM, Peutherer JF: Does infection with HIV affect the outcome of pregnancy? Br Med J 1988; 296:467.
9. Robertson JR, Bucknall AB: Pregnancy and HTLVIII/LAV transmission in heroin users. Health Bull 1986; 44(6):364–366.
10. Smith PF, Mikl J, Truman BI, *et al:* HIV infection among women entering the New York State correctional system. Am J Public Health 1991; 81(Suppl):35–40, Ch VI.