

Backcalculation of the Number Infected with Human Immunodeficiency Virus in Germany

*Johannes Seydel, *†Alexander Krämer, ‡Philip S. Rosenberg, *Knut M. Wittkowski, and ‡Mitchell H. Gail

*Department of Medical Biometry, University of Tübingen, Tübingen, Germany, †Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, Minnesota, and ‡Epidemiologic Methods Section, Epidemiology and Biostatistics Program, National Cancer Institute, Rockville, Maryland, U.S.A.

Summary: The method of backcalculation was used to estimate the cumulative number of human immunodeficiency virus type 1 (HIV-1)-infected adults in Germany from reporting delay-corrected surveillance data on acquired immunodeficiency syndrome (AIDS) in the pretreatment era. Using different backcalculation approaches with various incubation period distributions, a plausible range of 13,100 to 23,900 HIV-1-infected adults as of December 31, 1984, was calculated. Estimates of the number infected at more recent times were subject to much greater uncertainty. On average, the cumulative incidences calculated by the nonparametric backprojection method are about 15% lower than the results from the step function model. Nonparametric backprojection with the Hessel incubation distribution suggests a declining HIV infection rate after 1985, as might be expected from German health policies. This distribution is derived from cohorts of homosexual men, the main fraction of German AIDS cases. **Key Words:** Backcalculation—Backprojection—HIV incidence—Epidemiologic methods.

The method of backcalculation was developed to give estimates of individuals already infected with the human immunodeficiency virus type 1 (HIV-1) in order to project short-term incidence of the acquired immunodeficiency syndrome (AIDS) (1-4). The method uses information about the AIDS incubation-period distribution, which characterizes the latency period between infection with HIV-1 and AIDS onset, to "backcalculate" prior rates of HIV-1 infection from the numbers of reported AIDS cases. Projections of future AIDS incidence can then be obtained by allocating the numbers of

persons previously infected forward in time, using rates of disease progression determined from the incubation distribution.

Effective therapies, such as zidovudine (5,6) and aerosolized pentamidine (7) have recently prolonged the incubation period in significant numbers of AIDS-free persons with advanced HIV-1 infection (8,9). The effects of therapy can distort estimates of HIV prevalence and future projections of AIDS incidence obtained from backcalculation, which, in its original form, was predicated on the assumption that the incubation distribution did not change in calendar time. For this reason, we present estimates of the number infected with HIV during the pretreatment era. Furthermore, we evaluate and compare the impact of different backcalculation methods and various incubation-period distributions on backcalculation results.

Address correspondence and reprint requests to Dr. A. Krämer, Institute of Medical Biometry, University of Tübingen, Westbahnhofstrasse 55, N2070 Tübingen, Germany.

Manuscript received January 21, 1992; accepted June 7, 1993.

METHODS

German AIDS Surveillance Data

Since 1982, patients with the diagnosis of AIDS are reported to the AIDS Center in the Federal Republic of Germany. The AIDS cases were diagnosed on the basis of case definitions and exposure groups provided by the Centers for Disease Control and World Health Organization. Counts of AIDS incidence were tabulated from data reported to the German AIDS Center before May 31, 1992. Until May 1991, 6,455 AIDS cases were reported (70.1% homosexual/bisexual men, 13.2% intravenous drug users, 6.5% blood product recipients, 4.2% heterosexual risk partners, 0.6% prenatal or perinatal infections, and 5.4 with unknown risk).

The series was adjusted for reporting delays in monthly intervals using the following method, which allows for changes in reporting behavior (10). Denote by y_{jk} the number of new cases diagnosed for period j ($0 \leq j \leq t$, where t is the most recent period) and reported with a delay of k periods. Under the assumption that no cases are reported with a delay of more than K periods, the total number of cases diagnosed for period j is $N_j = \sum_{k=1 \dots K} c_{jk}$, where $c_{jk} = E(y_{jk})$ for future periods ($j + k > t$), and $c_{jk} = y_{jk}$ elsewhere. The nonparametric maximum-likelihood methods of Brookmeyer and Damiano (3), Heisterkamp et al. (11) and Rosenberg (12) are equivalent to estimating recursively

$$E(y_{jk}) = \sum_{j' < j} c_{j'k} \sum_{k' < k} c_{j'k'} / \sum_{j' < j, k' < k} c_{j'k'}$$

To take into account that the delay distribution may change over time, and to give more reliable estimates for recent periods, this estimate can be modified toward constrained moving base maximum likelihood (CMBML) estimates based on the previous b periods (months) only (10). The distribution of diagnosed cases is locally [i.e., for the following $K - k$ delay periods and the previous $b + (K - k)$ diagnosis periods] approximated by geometrical increase or decrease (i.e., $N_{t'}/N_{t'-1} = r$)

$$E_b(y_{jk}) = \frac{\sum_{j-b < j' < j} c_{j'k} \left(\sum_{k' < k} c_{j'k'} + \sum_{k' > k} c_{j-(k'-k),k} \right)}{\sum_{j-b < j' < j} \left(\sum_{k' < k} c_{j'k'} + \sum_{k' > k} c_{j'-(k'-k),k} \right)}$$

The base width b is chosen to minimize the χ^2 prediction error over the previous 12 months. For the German data, the optimal base width was 10 months.

A possibility of reporting delays of up to 5 years was allowed. This would have resulted in slightly higher adjusted estimates of AIDS incidence than those obtained by the method of Rosenberg (12). However, our estimates were in fact lower because of improvements in reporting behavior primarily in German metropolitan cities. The reporting delay adjusted series were calculated for persons ≥ 14 years of age only, because younger persons have a substantially different natural history (13). Only AIDS diagnoses through 1987 were used for backcalculation for the HIV-1 infection curve, because treatment may have modified the incubation period after that time (8,9).

Statistical Models

Detailed descriptions with further references for the used backprojection methods are given in Rosenberg et al. (14) for the

flexible parametric approach with step function models and Becker et al. (4) for the nonparametric approach.

In brief, the infection curve $v(s)$ specifies the instantaneous infection rate as a function of calendar time. Cumulative HIV infections and projections for future AIDS incidence are readily calculated from the infection curve. Assuming that the epidemic began in 1977, the expected AIDS incidence $E(Y_j)$ occurring in calendar quarter $[T_{j-1}, T_j]$ is linked to the infection curve.

$$E(Y_j) = \int_{1977}^{T_j} v(s)[F(T_j - s) - F(T_{j-1} - s)] ds,$$

where $F(t)$, the incubation period distribution, is assumed to be known from studies of natural history. Using the observed and reporting delay adjusted AIDS incidence Y_j and an estimate of $F(t)$, the method of backcalculation "solves" this equation for $v(s)$. Methods for selecting and fitting step function models for $v(s)$ are given in Rosenberg et al. (14). The second backprojection approach uses a modification of an expectation maximization (EM) algorithm for maximum likelihood estimation with a smoothing step (EMS) as described in Becker et al. (4). The EMS method works without parametric assumptions about the form of the HIV infection curve.

We used the following incubation-period distributions, $F(t)$: (a) the Weibull curve $F(t) = 1 - \exp(-0.0021t^{2.516})$. This estimate was obtained from individuals from the National Cancer Institute's multicenter hemophilia cohort who were older than 20 years at time of infection (15). The median time-to-AIDS was estimated to be 10.03 years for this cohort; (b) an incubation period distribution derived from a cohort of homosexual and bisexual men who were recruited for hepatitis B vaccine trials (16); (c) distributions based on AIDS incidence and seroconversion rates in San Francisco's gay community (17,18).

RESULTS

Using backcalculation methods with step function and EMS approach with different incubation-period distributions, we estimated that between 13,100 and 23,900 adults in Germany had been infected with HIV-1 prior to December 31, 1984 (Table 1). On average, the cumulative incidences calculated by the EMS backprojection were about 15% lower than the results delivered by the step function model. The absolute differences were between

TABLE 1. Cumulative number^a of HIV-1 infected adults^b in the Federal Republic of Germany as of December 31, 1984

Distribution	Backprojection procedure	
	Step function model	EMS backprojection
Hessol	15,400	13,100
Bacchetti & Moss	22,200	18,600
Brookmeyer & Goedert	23,900	20,600

^a Estimates inflated by 10% for underreporting.

^b Excludes persons younger than 14 years.

2,300 and 3,600 adults, depending on the model used.

Figure 1 shows the estimated incidence of HIV-1 infections in Germany that we obtained by using the unsmoothed step function backprojection with various incubation-period distributions. According to this picture, infections increased in the years 1980 and 1981 and again in 1983, remaining at that level until 1987. Infection curves estimated from Brookmeyer and Goedert and Bacchetti and Moss incubation-time distributions revealed an unrealistic fluctuation of HIV infections. The Brookmeyer and Goedert distribution shows no new infections between July 1982 and March 1983, which seems quite implausible.

Figure 2 displays smoothed backprojection estimates with the EMS algorithm and different incubation-period distributions. The HIV-1 incidence begins to rise in 1981, reaching a peak in 1984. In the second half of 1984, there is a decline in HIV incidence, followed by a second peak in 1985. Although the Bacchetti and Moss distribution reveals a more rapid increase and a higher first maximum point than the Hessol curve, both show the same second peak.

This pattern is not seen for the Brookmeyer and Goedert incubation-period distribution (Weibull), which shows only a single maximum point at the

beginning of 1984, followed by a continuous decline of HIV incidence. In addition, the Weibull distribution yields a faster rise in HIV incidence than do the backprojections with the Hessol and the Bacchetti and Moss distributions.

DISCUSSION

The backcalculated range of 13,100–23,900 for the cumulative number of HIV-1-infected persons in Germany includes uncertainty about the backcalculation method and the incubation-period distribution. A more exact calculation of prevalence would subtract from the number infected those who had died (i.e., 100 of 200 adult AIDS cases as of December 1984, and 3,200 of 6,400 cases as of May 1991) as well as the number of intravenous drug users who died from causes unrelated to AIDS.

The uncertainties shown do not allow for random error, for which a bootstrap analysis would be useful (14). However, previous work with step function models indicates that this component of uncertainty is smaller than that from choice of the incubation-period distribution (19). For EMS backcalculation, it seems that stochastic uncertainty may be substantial, especially in the recent past (20).

In principle, the optimal strategy for estimating HIV-1 prevalence would be direct serologic testing

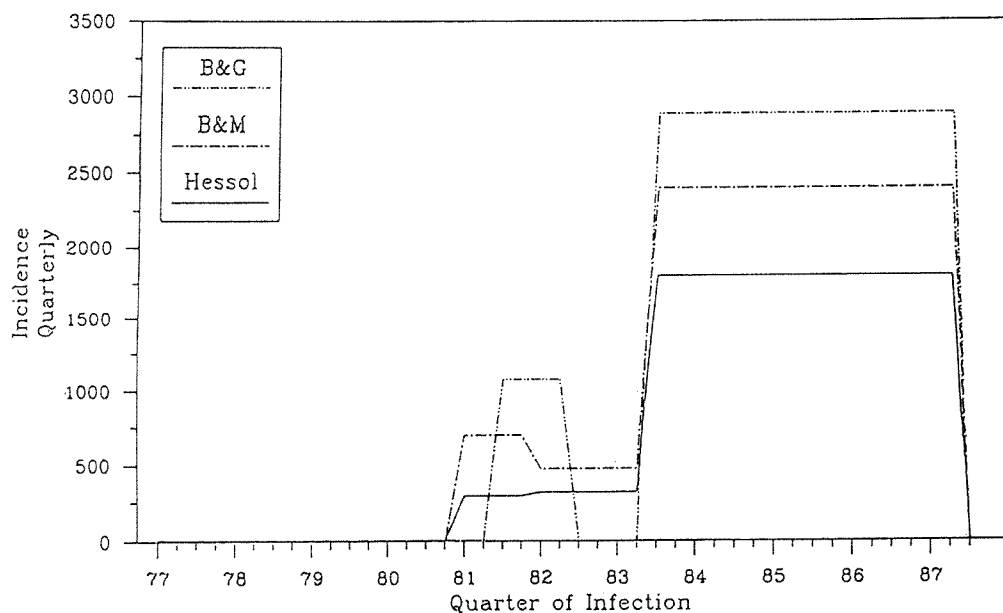


FIG. 1. HIV-1 incidence in the Federal Republic of Germany with the step function model. The quarterly incidence was estimated from observed and reporting delay-corrected AIDS incidence, using the three different incubation-time distribution models of Hessol, Bacchetti and Moss (B&M), and Brookmeyer and Goedert (B&G). A rise of infections was observed in the year 1981, with a plateau from 1983 to 1987. Note that the Bacchetti and Moss and the Brookmeyer and Goedert incubation models gave an unrealistic fluctuation, with no new infections between July 1982 and March 1983 for the Brookmeyer and Goedert distribution.

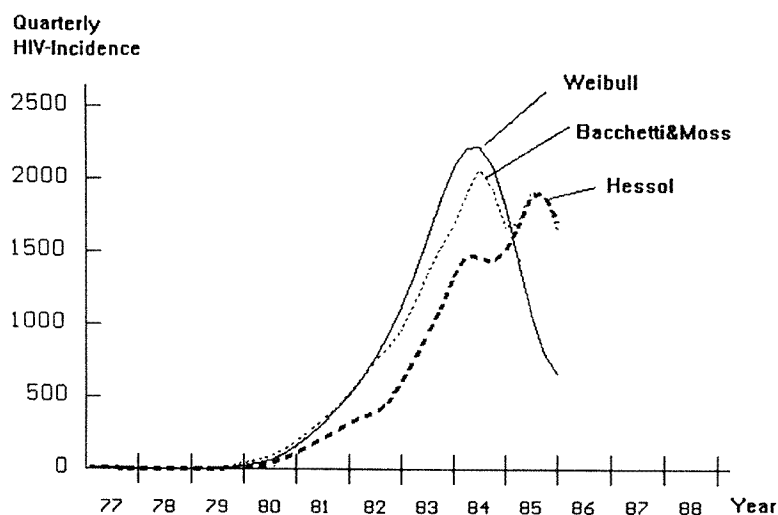


FIG. 2. HIV-1 incidence in the Federal Republic of Germany with the smoothed EMS backprojection approach and various incubation-period distributions. After the initial rise in 1981, the Weibull distribution shows a single peak in the first quarter of 1984 followed by a continuous decline of HIV incidence. The other distribution models reveal a second maximum point in 1985 and a slower increase of HIV incidence in the years before.

of a probability sample of the general population. However, nonresponse bias is a significant practical problem with this approach (21). Another strategy used in the United States employs the method of anonymous unlinked testing of certain groups of subjects, such as hospital applicants, Job Corps applicants, civilian applicants to the military, and childbearing women. Such data are not available in Germany. The method of backcalculation yields very uncertain estimates of the number of persons infected through June 30, 1987 or later, because of inherent limitations of the backcalculation method (19). We doubt that our estimates of HIV prevalence have been influenced by the effects of treatment, because we used incidence data only through June 30, 1987. Backcalculation tells us very little about the numbers infected from end 1984 to mid 1987. However, step function models with long last steps are likely to overestimate the recent infection rate if the infection curve is sharply decreasing (22). It is possible that backcalculation with EMS is biased, too, but necessary simulation studies remain to be done. The main source of uncertainties in estimates through December 31, 1984, is choice of incubation distribution. Exactly how the infection curve is modeled has a comparatively minor effect.

The preceding figures of the cumulative number of HIV-1-infected persons are in agreement with estimates reported from the German AIDS Center of the total number of HIV-1-infected persons in Germany (23). It is considered unlikely that further rapid epidemic spread of HIV-1 infections has occurred in Germany in recent years because there has been no increase in other sexually transmitted

diseases. We therefore assume that the number of HIV-1-infected persons in this country currently does not exceed 30,000 (24).

The unsmoothed backprojection with the step function shows an initial peak of HIV-1 incidence in the years 1981 through 1982, which is followed by a sharp increase in 1983 and a plateau until 1987. The Brookmeyer and Goedert (Weibull) distribution shows a quite implausible incidence decrease to zero infected persons in 1982 through 1983.

Using the smoothed EM backcalculation approach, estimates of HIV-1 incidence in Germany revealed a rise, starting in the year 1980, an initial peak of new infections in the first quarter of 1984, and apart from the Weibull distribution, a second peak in 1985 (Fig. 2).

The increase of the quarterly numbers and the magnitude of the peaks varies depending on the incubation-period distributions used. The Brookmeyer and Goedert (Weibull) and Bacchetti and Moss distributions show a fast rise and an early well-marked peak of HIV-1 incidences in the first quarter of 1984. The Hessol distribution delivers a slower increase and a maximum point in 1985. The Bacchetti and Jewell incubation-period distribution (18), which is also based on cohorts of gay men, gave a shape very similar to the Hessol distribution.

This general pattern of new infections is consistent with the concept that the epidemic evolved in Germany with a delay of about 2 years after that in the United States. Public health policies, such as blood-donor screening for HIV, heat treatment of blood factor concentrates, information campaigns, and behavioral changes make a decline in HIV in-

idence after a maximum in 1985 plausible, as it is shown by the Hessol incubation-period distribution, using the EMS backcalculation algorithm.

The Hessol cohort consisted of homosexual men who were recruited in the San Francisco area in the late 1970s for the purpose of hepatitis B vaccine trials. Because ~70% of the German AIDS cases were homosexual men, it may be appropriate to use an incubation-period distribution for backcalculation that is derived from homosexual men rather than from hemophilic persons, who account for only a small proportion of AIDS cases in Germany (4.5%). Differences between exposure groups with respect to the incubation-period distribution appear quite likely if one considers, for example, that Kaposi's sarcoma is much more frequent in homosexual men than in hemophilic persons and that Kaposi's sarcoma tends to occur earlier than opportunistic infections (25).

Acknowledgment: We thank Professor Dr. M. A. Koch who kindly provided us with AIDS incidence figures reported to the German AIDS Center, Bundesgesundheitsamt, Berlin, Germany. In addition, we are grateful to Laurel Decher, Division of Epidemiology, University of Minnesota, Minneapolis, Minnesota, U.S.A., Dr. Robert J. Biggar, Viral Epidemiology Branch, National Cancer Institute, U.S.A., and Professor Dr. K. Dietz, Department of Medical Biometry, University of Tübingen, Germany, for their fruitful suggestions.

REFERENCES

- Brookmeyer R, Gail MH. Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet* 1986;2:1320-2.
- Brookmeyer R, Gail MH. A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J Am Stat Assoc* 1988;83:301-8.
- Brookmeyer R, Damiano A. Statistical methods for short-term projections of AIDS incidence. *Stat Med* 1989;8:23-4.
- Becker NG, Watson LF, Carlin JB. A method of nonparametric backprojection and its application to AIDS data. *Stat Med* 1991;10:1527-42.
- Fischl MA, Richman DD, Grieco MH, et al. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex. *N Engl J Med* 1987;317:185-91.
- Volberding PA, Lagakos SW, Koch MA, et al. Zidovudine in asymptomatic human immunodeficiency virus infection: a controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. *N Engl J Med* 1990;322:941-9.
- Golden JA, Chernoff D, Hollander H, et al. Prevention of *Pneumocystis carinii* pneumonia by inhaled pentamidine. *Lancet* 1989;1:654-7.
- Gail MH, Rosenberg PS, Goedert JJ. Therapy may explain recent deficits in AIDS incidence. *J Acquir Immune Defic Syndr* 1990;3:296-306.
- Rosenberg PS, Gail MH, Schragger LK, et al. National AIDS incidence trends and the extent of zidovudine therapy in selected demographic and transmission groups. *J Acquir Immune Defic Syndr* 1991;4:392-401.
- Wittkowski KM, Dietz K. Backcalculation methods and their application for estimating HIV prevalence in Germany. *Bulletin of the International Statistical Institute—Contributed Papers*, vol. 2. 48th I.S.I. Session, Cairo, Egypt, Sept. 9-17, 1991:710-1.
- Heisterkamp SH, Jager JC, Downs AM, et al. Correcting reported AIDS incidence: a statistical approach. *Stat Med* 1989;8:963-76.
- Rosenberg PS. A simple correction of AIDS surveillance data for reporting delays. *J Acquir Immune Defic Syndr* 1990;3:49-54.
- Goedert JJ, Kessler CM, Aledort LM, et al. A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N Engl J Med* 1989;321:1141-8.
- Rosenberg PS, Biggar RJ, Goedert JJ, Gail MH. Backcalculation of the number with human immunodeficiency virus infection in the United States. *Am J Epidemiol* 1991;133:276-85.
- Brookmeyer R, Goedert JJ. Censoring in an epidemic with application to hemophilia-associated AIDS. *Biometrics* 1989;45:325-35.
- Hessol NA, Lifson AR, O'Malley PM, et al. Prevalence, incidence, and progression of human immunodeficiency virus infection in homosexual and bisexual men in hepatitis B vaccine trials, 1978-1988. *Am J Epidemiol* 1989;130:1167-75.
- Bacchetti P, Moss AR. Incubation period for AIDS in San Francisco. *Nature* 1989;338:251-3.
- Bacchetti P, Jewell NP. Nonparametric estimation of the incubation period of AIDS based on a prevalent cohort with unknown infection times. *Biometrics* 1991;47:947-60.
- Rosenberg PS, Gail MH. Uncertainty in estimates of HIV prevalence derived by backcalculation. *Ann Epidemiol* 1990;1:105-15.
- Becker NG, Watson LF. Use of empirical transformations in nonparametric back-projection of AIDS incidence data. In: Jewell NP, Dietz K, Farewell VT, eds. *AIDS epidemiology: methodological issues*. Boston: Birkhäuser-Boston, 1992:1-38.
- CDC. Pilot study of a household survey to determine HIV seroprevalence. *MMWR* 1991;40:1-5.
- Rosenberg PS, Gail MH, Pee D. Mean square error of estimates of HIV prevalence and short-term AIDS projections derived by backcalculation. *Stat Med* 1991;10:1167-80.
- AIDS-Zentrum des Bundesgesundheitsamtes. *AIDS-Nachrichten Forsch Wissensch* 1990;2:1-16.
- Wittkowski KM, Dietz K, Koch MA. Einfluss von Modellannahmen auf die Schätzung der HIV-Prävalenz. In: Hartmann U, Grutza C, eds. 3. *Deutscher AIDS-Kongress*. Munich: Futuramed-Verlag, 1990:469.
- Lifson AR, Darrow, Hessol NA, et al. Kaposi's sarcoma in a cohort of homosexual and bisexual men: epidemiology and analysis for cofactors. *Am J Epidemiol* 1990;131:221-31.