Journal of Acquired Immune Deficiency Syndromes
6:1150–1156 © 1993 Raven Press, Ltd., New York

Survival Differences and Trends in Patients with AIDS in the United States

Xin Ming Tu, *Xiao-Li Meng, and †Marcello Pagano

*Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania; *Department of Statistics, University of Chicago, Chicago, Illinois; and †Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, U.S.A.*

Summary: The AIDS surveillance system maintained by the Centers for Disease Control and Prevention (CDC) provides a unique data base for estimating survival after a diagnosis of AIDS for the general AIDS population in the United States. Because patients enrolled in most AIDS clinical trial studies receive unusual medical care that may not be available to the general public and typically have relatively longer survival time, estimates obtained from these studies may not be of direct use in assessing the national health-care needs. Furthermore, such studies are usually of short duration and may not be very informative for long-term health-policy planning. We present survival estimates obtained from the CDC surveillance data for the adult/adolescent AIDS population in the United States and compare their survival and trend in survival on gender, sexual behavior, and injection-drug use status. These estimates provide information for mortality risk after an AIDS diagnosis over a period of 8 years and for trend of survival during the period between 1983 and 1991. **Key Words:** CDC surveillance—Homosexual males—Injection drug use—Pneumocystis carinii pneumonia—Proportional hazards—Survival analysis.

Long-term health care and policy planning for AIDS require estimates of risk of mortality. Inpatient and clinical trial studies (e.g., 1,2) are extremely important for medical and clinical research, but are probably less so for health departments because patients in these studies receive unusual medical care that may not be available to the general public and they tend to survive longer (e.g., 3). The AIDS surveillance system maintained by the Centers for Disease Control and Prevention (CDC) on the other hand provides a data source that reflects the general situation of the AIDS population in the United States. There are also several other features of the CDC data base that are not seen elsewhere. First, data from the CDC are collected almost over the entire history of the AIDS epidemic. Survival estimates can then be obtained to reflect temporal trend over the years. Second, this is the only data base that is close to covering the entire AIDS population in the United States. Survival differences of epidemiologic interest can be examined at the national level.

Because of the useful information contained in the CDC AIDS surveillance system, it has long been a goal to utilize this data base for estimating the risk of mortality for patients with AIDS (e.g., 4–7). However, due to various analysis difficulties, estimates from these earlier analyses are either severely biased (e.g., 4) or confined to a particular group of AIDS patients that may not be representative of the AIDS population in general (5–7). In this article, we present estimates of survival for the adult/adolescent AIDS population in the United

Address correspondence and reprint requests to Dr. X. M. Tu at Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA, 15260, U.S.A.

Manuscript received July 26, 1992; accepted February 1, 1993.

States. These estimates are obtained from the deaths reported to the CDC surveillance system by utilizing some recent development in statistical methodology. They provide information for the risk of mortality over the first 8 years after an AIDS diagnosis and for the trend of survival during the period between 1983 and 1991. These estimates may be used to serve as a basis for estimating other statistics of epidemiologic interest such as AIDS prevalence (3).

DATA SOURCE AND STUDY POPULATION

The estimates presented here are based on the deaths of the adult/adolescent AIDS population (a few transfusion-related AIDS cases were not included) in the United States reported to the CDC as of July 1991, with an AIDS diagnosis between the first quarter of 1983 and the first quarter of 1991. Because of severe underreporting, deaths with AIDS diagnosed in the second quarter of 1991 were excluded. In addition to separating the male and female subpopulations in the analysis, the cases were further divided on the basis of different modes of exposure to HIV as well as different manifestations of AIDS (the mode of exposure indicated in the data base was used for patients with multiple risk exposures). For the male subpopulation, four major risk groups were considered on the basis of their sexual behavior and injecting drug (ID) use status: men who have sex with men (including bisexual contact) with ID use (6,329 deaths) and non-ID use (60,530), denoted IDMSM and MSM, respectively; and heterosexual men with ID use (14,533) and non-ID use (847), denoted IDMSW and MSW, respectively. The female subpopulation was divided into two subgroups: one with ID use (4,139) and one without ID use (2,427), denoted by IDWSM and WSM, respectively.

Within each of these risk groups, three diagnosis strata were defined in the following order: *Pneumocystis carinii* pneumonia (PCP), disease manifestations other than PCP and Kaposi's sarcoma (OTH), and Kaposi's sarcoma (KS). Note that both definitive and presumptive diagnoses were used in creating the strata and a similar classification scheme was also used by Lemp et al. (6). Individuals with multiple diagnoses were classified according to the highest ranked stratum. For example, individuals diagnosed with PCP and other opportunistic infection(s) were classified into the PCP category, those diagnosed with other opportunistic infection(s) and KS were classified into the OTH category, etc. Thus, under this classification scheme, a patient in the PCP category only indicates that the patient had PCP as one of the diagnoses, which may not have been the first one diagnosed. Note that to minimize the influence of the change of definition, patients with a disease added to the case definition in 1987 (basically wasting, dementia, and disseminated TB; see ref. 8) were excluded from the analysis.

METHODS

Analysis of survival time based on the CDC AIDS surveillance data is complicated by the fact that not all deaths are reported. Unlike the studies in Harris (5) and Lemp et al. (6,7), where patients had been closely followed up for vital information such as death, there is a sizable fraction of reported AIDS patients whose deaths will never be reported to CDC. As a result, analysis based on reported AIDS cases in the CDC data base using their approach leads to severe upward bias (see Discussion). One way to avoid this problem is to condition the analysis on the deaths reported to the CDC surveillance system. This approach requires several statistical considerations, among which a basic problem is to adjust the underreporting of deaths caused by delays in reporting. By utilizing the information on death reporting recently added to the CDC surveillance data base, methods for proper survival analysis under this approach have been recently developed (9) by using the EM algorithm (10) and multiple imputation (11). They provide the methodologic basis for the estimates presented here.

Prior to the analysis, the times of the observed events of interest (AIDS diagnosis, death, reporting of death) were grouped into intervals of a quarter of a year. This discretization scheme was used because both death and reporting of death were given in calendar quarters (8).

The statistical model used in modeling survival is the discrete proportional hazards model with the probability mass function given by

$$f(j|z) = \begin{cases} (p_0 \cdots p_{j-1})^{\exp\{z^T\beta\}}(1 - p_j^{\exp\{z^T\beta\}}) & \text{if } 0 \leq j \leq J-1 \\ (p_0 \cdots p_{J-1})^{\exp\{z^T\beta\}} & \text{if } j = J \end{cases}$$

Here, $f(j|z)$ is the probability that a death occurs during quarter j after an AIDS diagnosis, with $j = 0$ denoting the quarter of death being the same as the quarter of diagnosis: z denotes a covariate vector that can consist of ordinal variables for time of AIDS diagnosis, indicators for risk categories, etc.; p_j $(0 \leq j \leq J-1)$ and β is a vector of unknown parameters. This model was chosen in preference to several other alternatives mainly because it provides a similar interpretation of model parameters as the continuous proportional hazards model, which is widely used in epidemiologic studies. Like the continuous proportional hazards model, the hazard $h(s|z)$ here is also approximately "proportional":

$$h(j|z) \approx \exp\{a_j + z^T\beta\},$$

where $a_j = \log[-\log(p_j)]$ is a constant. A consideration in choosing the discrete model as opposed to the continuous one in the current context is the stability of the estimators in such models when used to analyze data with numerous ties (as in this dataset).

The reporting delay distributions were also modeled using the discrete proportional hazards model with covariates accommodating both the temporal trend over the time of death and variation among the five geographic regions consisting of metropolitan statistical areas with population at least 1 million: Northeast, Central, West, South, Mid-Atlantic, and a residual category that consists of areas with population <1 million. All deaths from the population considered that occurred between January 1986 and March 1991 were included in the analysis. The missing dates of death reporting for cases diagnosed before October 1987 were treated left-censored in the analysis using the EM algorithm.

To use these estimates for correcting for the underreporting of deaths in survival analysis, we assume that the number of deaths not reported within 5 years is negligible. We checked this assumption against the data and found it to be very reasonable. The survival distributions were estimated by adjusting the reported deaths for underreporting using the estimates of the reporting

delay distributions. Multiple imputation was used to incorporate both the sampling variation and variation in the estimates of the delay distributions (9).

RESULTS

Plotted in Figs. 1 to 3 are the estimated survival distributions for the various risk groups obtained using the discrete proportional hazards model with no time trend, which correspond to the survival distributions averaged over the period between 1983 and 1991. These distributions reflect survival over a period of 8 years following diagnosis. In all cases, the ID groups show a relatively shorter survival time than the non-ID risk groups. For PCP cases, Fig. 1 shows that MSM have the longest survival (survival median is ~15 months for non-ID and ~12 months for ID users), and the male heterosexuals have a similar survival distribution as the female heterosexual population (survival median is about 12 months for non-ID and 9 months for ID users). Inspection of the survival curves in Fig. 2 indicates that there is not much difference in survival among MSM, male heterosexuals, and female heterosexuals in this disease category (after controlling for ID drug use status): the survival medians are ~9 months for the non-ID and ~7 months for the ID users. As shown in Fig. 3, the survival median for MSM with KS is about 15 months. Note that other risk groups in this category were not included because of the small numbers of such cases. The survival estimates in Figs. 1–3 seem to be a little lower than those reported in Harris (5), in Lemp et al. (6) and in Friedland et al. (2) for the corresponding risk groups. The difference may well be due to the selection bias when confined to the cohorts in these studies. The dip at the end of the third month for PCP (Fig. 1), which was also observed in Harris (5) and in Friedland et al. (2), seems to suggest that patients with PCP as one of the diagnoses might be at a relatively high risk during the first 3-month period following diagnosis.

The survival trend for each disease and risk-group category was examined by fitting the model with a covariate coded 1–9, designating the year of diagnosis between 1983 and 1991. Most of the groups considered show some improvement in survival during this period. However, at the 1% level, significant improvement in survival is only confined to the MSM population: ~13% annual reduction rate in mortality for both ID and non-ID users with PCP and ~7% annual reduction rate for non-ID with OTH in this population (all with p values of <0.00001).

Plotted in Fig. 4 are the quartiles and medians of the three survival distributions that show significant

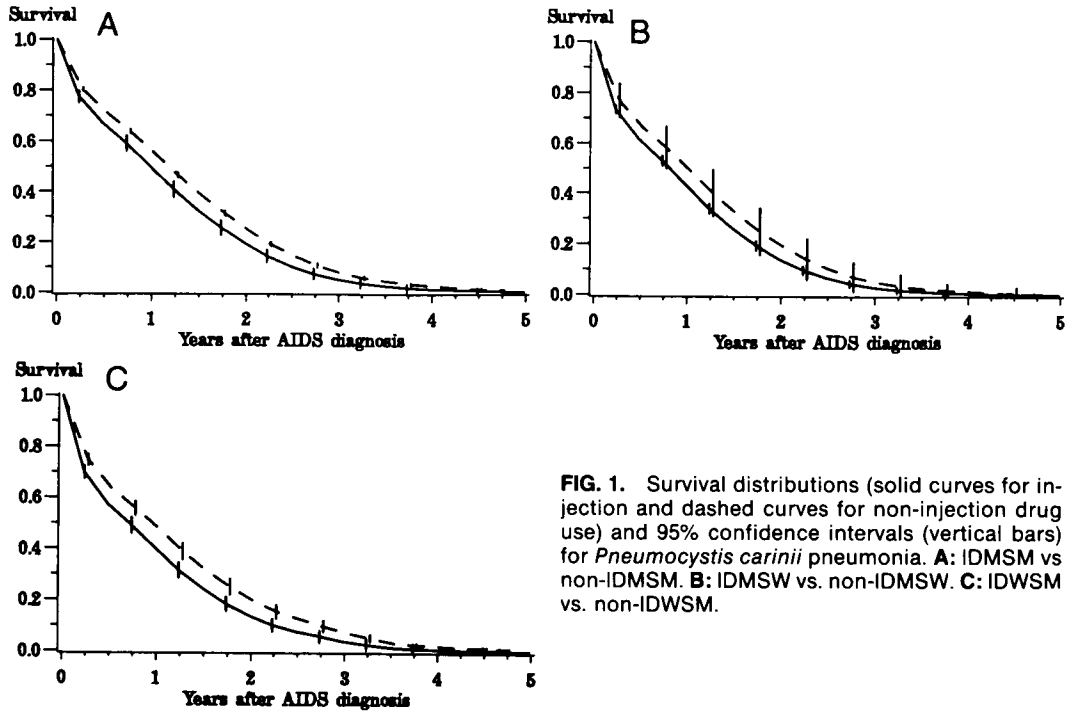

FIG. 1. Survival distributions (solid curves for injection and dashed curves for non-injection drug use) and 95% confidence intervals (vertical bars) for *Pneumocystis carinii* pneumonia. **A:** IDMSM vs non-IDMSM. **B:** IDMSW vs. non-IDMSW. **C:** IDWSM vs. non-IDWSM.

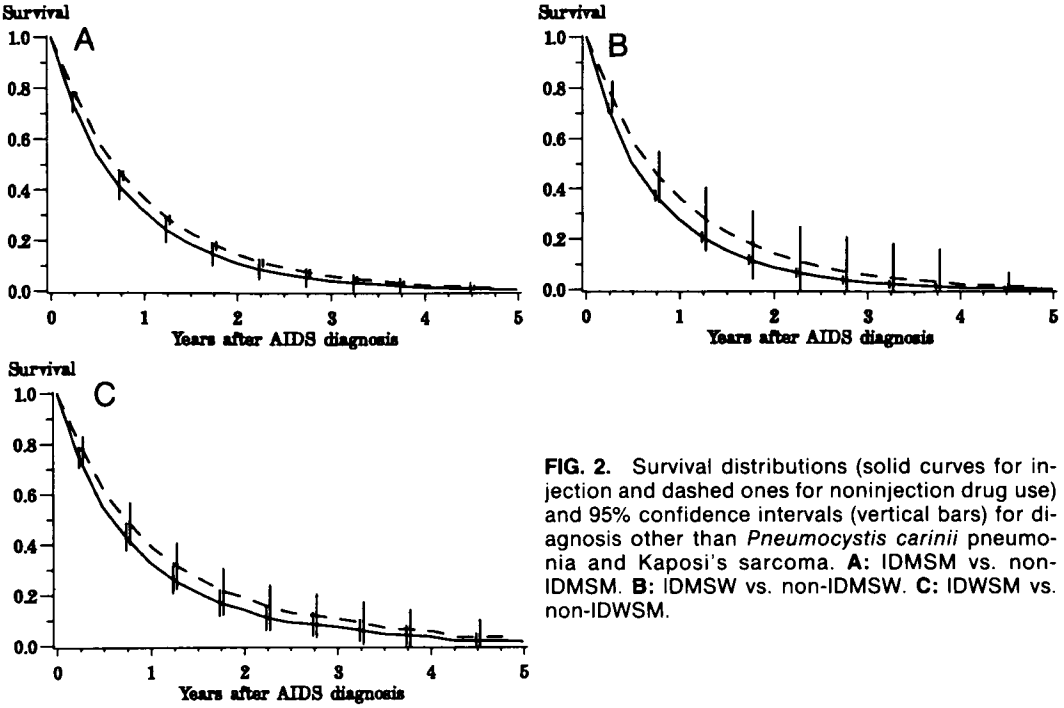

FIG. 2. Survival distributions (solid curves for injection and dashed ones for noninjection drug use) and 95% confidence intervals (vertical bars) for diagnosis other than *Pneumocystis carinii* pneumonia and Kaposi's sarcoma. **A:** IDMSM vs. non-IDMSM. **B:** IDMSW vs. non-IDMSW. **C:** IDWSM vs. non-IDWSM.

improvement in survival between 1983 and 1991. In particular, the survival medians for the non-ID MSM with PCP averaged over the periods between 1983–1985, 1986–1987, and 1988–1990 are ~10, ~15, and ~21 months, respectively, which are slightly lower than the medians of 10.3, 17.9, and 21 months reported for the PCP cases diagnosed in the periods 1981–1985, 1986–1987, and 1988–1990 in the San Francisco cohort (Lemp et al., 7). Note that only the estimates for the non-ID users are comparable to the estimates from the San Francisco study, which may not be surprising because ~85% of the population in the San Francisco study were non-ID homosexual men. The survival medians for non-ID MSM with OTH averaged over the periods 1983–1985, 1986–1987, and 1988–1990 are ~7, 8 and 10 months, which are all lower than the average medians of 9, 10, and 13 months over the periods 1981–1985, 1986–1987, and 1988–1990 for the same disease category reported in Lemp et al. (7). As noted earlier, the difference between the estimates may be attributed to the difference in survival between the San Francisco cohort and the AIDS patients in general.

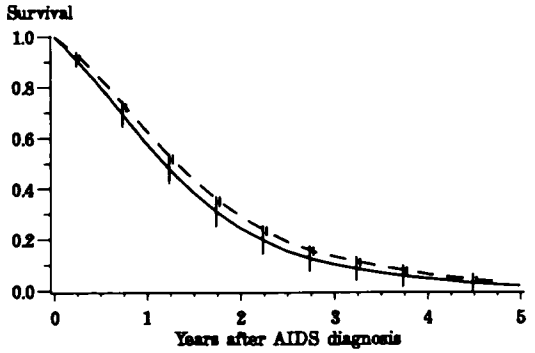

FIG. 3. Survival distributions (solid curve for injection and dashed one for noninjection drug use) and 95% confidence intervals (vertical bars) for Kaposi's sarcoma for IDMSM vs. non-IDMSM.

Improvement of survival for PCP after 1987 has been reported by short-term survival analysis (5,7). To accentuate the difference in survival between the two periods, we also fitted models with a binary covariate (1 for diagnosis after 1987 and 0 otherwise) to the PCP cases in the MSM risk group. The estimated reduction rates in mortality after 1987 for the ID and non-ID groups are, respectively, 28% and 40%, which seem to be comparable to the 30% found in the San Francisco study for patients taking AZT (6,12) and some other clinical trial studies (1). Based on these other studies, it has been hypothesized that the improvement for PCP after 1987 was partly due to the intervention of the antiviral therapy AZT (5), though improvement of survival is evident even before 1987, a fact that must also be

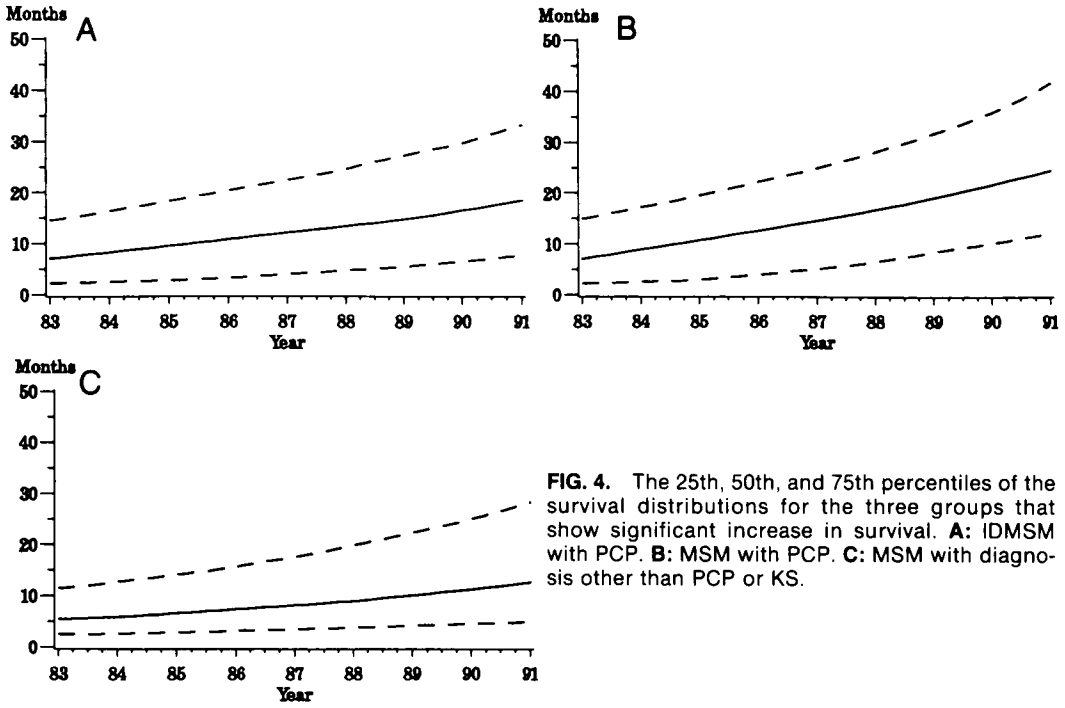

FIG. 4. The 25th, 50th, and 75th percentiles of the survival distributions for the three groups that show significant increase in survival. **A:** IDMSM with PCP. **B:** MSM with PCP. **C:** MSM with diagnosis other than PCP or KS.

part of any explanation that attributes the improvement to some key event that occurred in 1987 (13,14). Note that the reduction in mortality for non-ID users seems to be slightly higher than that reported by the San Francisco study, which might suggest that the improvement of survival for the PCP cases could also be attributed to other treatments such as trimethoprim-sulfamethoxazole or aerosolized pentamidine that have been found to be successful when given prophylactically in preventing recurrence of PCP (15,16). Of course, because of the observational nature of the surveillance data, the estimates presented here do not provide confirmatory evidence for these hypotheses.

DISCUSSION

The estimates reported here reflect survival after a diagnosis of AIDS, not after a contraction of AIDS. These survival estimates are probably more relevant when the surveillance data base is used for the purposes of assessing health-care needs and long-term policy planning, though they may not provide much information for some biologic issues such as the timing of intervention of a treatment.

In the analysis, deaths occurring within the same quarter as diagnosis were treated as having survival times of <3 months. However, a proportion of these individuals may represent a delayed diagnosis of AIDS (perhaps even at death) rather than rapid progression of disease (12). This may be especially true for the MSW, who showed a slight decline in survival between 1983 and 1991 (the p values for the various disease categories are in the range of 0.7). It is difficult or even impossible to identify these individuals from the surveillance data. As an attempt to investigate this possible bias, we refitted the models to the MSW after deleting the deaths that occurred within the same quarter as diagnosis, and there was very little change in the estimates.

We also examined the deaths that occurred before 1984 and found that there was <1% of them who survived for 8 years. In light of this and results from other AIDS surveillance data and AIDS clinical trials studies, we think that it would be rare for an individual to survive for >8 years after an AIDS diagnosis. The estimates reported here therefore provide good approximations to the survival distributions even though they are obtained by conditioning on death within 8 years.

The proportional hazards model fits reasonably well to the data as assessed by the G^2 statistic (e.g., 17). The p values calculated under this statistic for the models fitted to the various risk groups vary from 0.24 to 0.98. The assessment for linear trend was based on the comparison between the models with linear time trend and the ones with indicators

for the year of diagnosis. For the risk groups that exhibit significant time trend in survival, comparisons of goodness of fit using the G^2 statistic do not seem to show much improvement for the models with the indicators for the year of diagnosis.

It is quite possible that the change in definition of AIDS by the CDC in 1987, which broadened the case definition to include presumptive diagnosis of several conditions, including PCP (8), early diagnosis, overall improvement of medical care and treatment, etc., could all affect the estimates. The results presented suggest that any explanations must accommodate a differential effect on the risk groups, because significant improvement in survival is not observed for the heterosexual population.

The survival estimates presented here are based on reported deaths rather than on reported AIDS cases using standard survival methodology as in several other analyses (5,6). Unlike these other studies where patients had been closely followed up after an AIDS diagnosis, there is a sizable fraction of deaths in the CDC data that will never be reported. Thus, the standard approach is not appropriate for analyzing the CDC data. Simply censoring reported AIDS cases with no death certificates at the time of analysis will cause severe bias. Figure 5 shows the comparison between the two approaches; the upper two survival curves are based on reported AIDS cases with death censored at two different times, and the bottom one, which is the same curve in Figure 3 for the same risk group, is based on reported deaths. The survival curves based on the reported AIDS cases show that 15–18% of PCP cases would still be alive at the end of 5 years, a survival rate that is much higher for PCP than those reported by most studies (2,6). This example explains the upward bias in the estimates reported by Rothenberg et al. (4) that was first recognized by Lemp et al. (6). Note that censoring death 1 year prior to the time of analysis produced a less biased estimate. However, such estimates from reported AIDS cases are always biased unless we have knowledge to separate the unreported deaths from those who were still alive at the time of analysis.

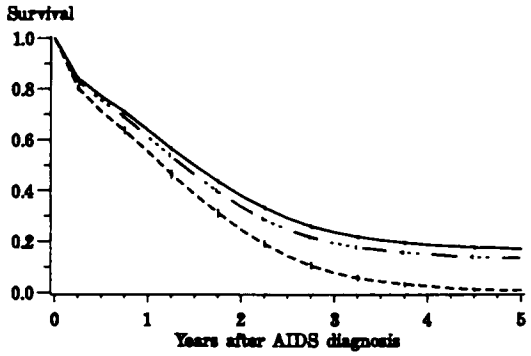

FIG. 5. Survival distributions and 95% confidence intervals (vertical bars) for the MSM risk group with *Pneumocystis carinii* pneumonia based on reported AIDS cases with death censored at the time of analysis, July 1991 (upper curve); reported AIDS cases with death censored after June 1990 (middle curve); and reported deaths (bottom curve).

Acknowledgment: We thank V. DeGruttola and R. Royce for helpful discussions, and two anonymous referees for constructive comments that led to the improvement of presentation. Research for X.M.T. and M.P. was supported in part by grants from the National Institutes of Health (NIAID-AI28076, T32-AI07358, R29-AI28905, NO1-AI-95030) and for X.-L.M. in part by NSF grant DMS-92-04504 and by University of Chicago/AMOCO fund.

REFERENCES

1. 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 millimeters. *N Engl J Med* 1990;322: 941–9.
2. Friedland GH, Saltzman B, Vileno J, et al. Survival differences in patients with AIDS. *J Acquir Immune Defic Syndr* 1991;4:144–53.
3. Pagano M, DeGruttola V, MaWhinney S, Tu XM. The HIV epidemic in New York City; statistical methods for projecting AIDS incidence and prevalence. In: Dietz K, Farewell V, Jewell NP, eds. *Statistical methodology for the study of the AIDS epidemic*. Boston: Birkhäuser-Boston, 1992; 123–40.
4. Rothenberg R, Woelfel M, Stoneburner R, et al. Survival with the acquired immunodeficiency syndrome. *N Engl J Med* 1987;317:1297–302.
5. Harris JE. Improved short-term survival among AIDS patients initially diagnosed with *Pneumocystis carinii* pneumonia, 1984 through 1987. *JAMA* 1990;263:397–401.
6. Lemp GF, Payne SF, Neal D, et al. Survival trends for patients with AIDS. *JAMA* 1990;263:402–6.
7. Lemp GF, Hirozawa AM, Araneta MR, Young K, Nieri G. Improved survival for persons with AIDS in San Francisco. In: VII International Conference on AIDS, Florence 1991. Abstract book: vol. 1. 1991:TU.C.41.
8. Centers for Disease Control. *AIDS public information data set*. Atlanta: CDC, 1991.
9. Tu XM, Meng XL, Pagano M. The AIDS epidemic: estimating survival after AIDS diagnosis from surveillance data. *J Am Statist Assoc* 1993;88:26–36.
10. Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm [with Discussion]. *J R Statist Soc B* 1977;39:1–38.

11. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley, 1987.
12. Centers for Disease Control. *MMWR* 1990;39:25–31.
13. Bennett CL, Garfinkel JB, Greenfield S, et al. The relation between hospital experience and in-hospital mortality for patients with AIDS-related PCP. *JAMA* 1989;261:2975–9.
14. Cotton DJ. Improving survival in acquired immunodeficiency syndrome: is experience everything? *JAMA* 1989;261: 3016–7.
15. Leoung GS, Feigal DW, Montgomery AB, et al. Aerosolized pentamidine for prophylaxis against *Pneumocystis carinii* pneumonia. *N Engl J Med* 1990;323:769–75.
16. Hirschel B, Lazzarin A, Chopard P, et al. A controlled study of inhaled pentamidine for primary prevention of *Pneumocystis carinii* pneumonia. *N Engl J Med* 1991;324:1079–83.
17. Agresti A. *Analysis of ordinal categorical data*. New York: Wiley, 1984.