INVITED

# A Bayes Application in HIV Infection Incorporating A Time Varying Hazard Model

Alfred A. Bartolucci, Karan P. Singh and Tonya M. Smoot

Department of Biostatistics, School of Public Health
Adolescent Medicine HIV/AIDS Research Network Data Center
The University of Alabama at Birmingham
Birmingham, Alabama 35294, U.S.A.
(Supported in part by grant NICHD U01 HD/GF32842)

**Summary** A statistical model is developed which accounts for the observed incidence of AIDS cases in the U.S. for a nine-year period. We derive the expected multinomial probability of developing AIDS within six-month time intervals from infection to AIDS. Information from the infection rate function is combined with these probabilities and averaged over the prior parameter space to model the overall cumulative distribution of time from infection to AIDS. A Gamma family of prior distributions is examined for modeling the system which accounts for changing risk over time with increasing hazard as one approaches the latter stages of infection. Thus, one incorporates a time varying hazard restriction on varying portions of time. Likelihood regions for the infection rate parameters are derived, and the rich family of prior distributions is fairly robust in modeling the cumulative time to AIDS for a number of years after seroconversion. The accuracy of the model for prediction is restricted to only a few time intervals beyond the end of the observed time period.

## 1. Introduction

Time to AIDS from seroconversion is distributed such that risk increases as one moves from the initial infection. The modeling of the size of the epidemic is well established in Brookmeyer and Gail (1988), Taylor (1989), Rosenberg, Gail and Pee (1991), Rosenberg, Bigger, Goedert and Gail (1991) and Anderson, Medley, May and Johnson (1986). Methodologies have included development of deterministic models which account in part for the many factors which affect the spread and development of AIDS. Other approaches include the development of stochastic models. We take this latter approach in furthering the methodology of Taylor (1989) in a Bayesian framework incorporating time varying hazard as is outlined in Blackstone, Naftel and Turner (1986). The assumption is that the hazard function of the time-related event to AIDS from the initial infection is time varying in a structured fashion as is the influence of risk factors associated with the events. In our current development, we assume a constant covariate structure or basically ignore any covariate influence on the change in hazard. Various phases are noted when the hazard does change over time. The Bayesian model provides a good fit to the data. For demonstration purposes we concentrate on the nine-year data from the first half of 1978 to the first half of 1987. Taylor (1989) uses this cohort, and it provides a good basis for our methodology in modeling the cumulative distribution of the time from infection to AIDS.

### 1.1 Statistical Methods

Let $I_i$ denote the incidence of HIV infections at time point i where we consider six-month increments, i.e. i = 1 corresponds to July 1978, i = 2 corresponds to January 1979, ..., i = 18 corresponds to January 1987. Let the incidence of HIV at time point i be

$$I_i(\lambda,b,c) = ce^{-\lambda e^{bi}}. \qquad (1)$$

This is the double exponential model with unknown parameters $\lambda$, b and c for $\lambda$, c > 0, b real. This is one of the incidence distributions considered by Taylor (1989). This represents an initial exponential growth with the possibility that a plateau is reached later in the epidemic. As in Taylor (1989) let $P_i$ i = 1, ..., 18 denote the probability of developing AIDS within i six-month time periods of infection. So $P_2$ = probability of developing AIDS in the first year, $P_{18}$ = probability of developing AIDS within nine years of infection. The $P_i$'s are derived empirically from the existing data. See Taylor (1989). We let $X_j$, j = 1, ..., 18 denote the number of AIDS cases which developed in a six-month period j from j = 1, the second half of 1978 to j = 18, first half of 1987. We let N = 49,842 be the total number of AIDS cases which developed by the middle of 1987. The distribution of $X_j$ is multinomial (N: $M_1$, ..., $M_{18}$) where $M_j$ is the probability that one of the AIDS cases from the set of N cases occurs in interval J. That is

$$M_j = \frac{1}{N} \sum_{i=1}^{j} I_i(\lambda, b, c)\,(P_{j-1-i} - P_{j-1}). \qquad (2)$$

where $P_o = 0$. $I_i$ occurs at the start of the $i^{th}$ time period. This equation is in the usual convolution format discussed in backcalculation. See Rosenberg, Gail and Pee (1991).

In the spirit of Blackstone et al. (1986), we wish to model the prior for $\lambda$ using decompositon into phases of time-varying risk. This is accomplished by a parametric modeling system that was originally conceptualized by Blackstone et al. (1986),

in terms of the cumulative hazard function $\Lambda(\cdot)$. Multiple overlapping phases of risk are consider to be additive, with each phase individually shaped by a function of time and scaled by a function of concomitant information. In our case, we omit consideration of concomitant data and our time-varying hazard corresponding to risk periods $l$ for changes in the hazard in times 1 to 18 is

$$\Lambda(\lambda) = \sum_{l=1}^{R} \tau_l(\tau_l\lambda)^{\gamma_l - 1}/N(\lambda) \qquad (3)$$

for

$$N(\lambda) = \frac{\lambda}{\Gamma(\gamma_l)} \int_{t_l}^{\infty} (\lambda\tau_l)^{\gamma_l - 1} e^{-\lambda\tau_l}\, d\lambda$$

where R equal the number of periods in which the hazard changes from interval 1 to interval 18, $\tau_l$, $\gamma_l > 0$ are the prior hyperparameters and $t_l$ is the midpoint for risk period $l$. The first period ($l = 1$), for example, is from interval 1 to interval 6. See Table 1. This article presents a model for decomposition of $\Lambda(\lambda)$ and later posterior $\Lambda(\lambda|t)$ into as many as R = 5 phases or periods. Each phase is defined for $0 < t < \infty$ although the effects of each are more prominent at one time than another. Prior scale and shape parameters are assessed accordingly. See Birch and Bartolucci (1983), and Bartolucci, Katholi and Birch (1992) for a methodology of deriving initial estimates of the hyperparameters.

The Bayesian prior to posterior analyses, thus, allows us to compute the value of each $M_j$ averaged over the prior parameter space for $\lambda$ with a prior density, $p(\lambda|\tau_l, \gamma_l\, jR)$. That is let

$$P_{ji} = P_{j-1,\,i} - P_{j-1} \qquad then$$

$$E(M_j|\lambda, \tau_l, \gamma_l; R) = \frac{1}{N} \sum_{i=1}^{j} P_{ji} \int_0^{\infty} I_i(\lambda, b, c) p(\lambda|\tau_l, \gamma_l) d\lambda \qquad (4)$$

$$= \frac{\tau^{\gamma}}{N} \sum_{i=1}^{j} P_{ji}/(\tau + e^{bi})^{\gamma}$$

where

$$p(\lambda|\tau_l, \gamma_l) = \delta_R(l) \frac{\lambda}{\Gamma(\gamma_l)} (\lambda\tau_l)^{\gamma_l - 1} e^{-\lambda\tau_l}$$

for

$$\delta_R(l) = 1 \text{ for period } l$$
$$0 \text{ otherwise.}$$

Thus, the posterior expected AIDS incidence within interval j for hazard period $l$ is $NE(M_j|\lambda, \tau_l, \gamma_l; R)$. Parameters c and b are those which maximize the posterior expected incidence (4).

### 1.2 Results

The value of c which maximized the posterior expected likelihood was c = 25432. The expected likelihood region for b was -.06 to +.06 with a peak at b = 0.0005. Note in Table 1 that the prior and posterior hazards are increasing which is consistent with increased risk as one moves forward in time from seroconversion. The hazard updated by the data is tempered compared to the prior hazard. Note the five-risk regions for time varying hazard. Period $l = 1$ is the first three years, period $l = 2$ is 3.5 to 4.5 years, etc. Table 2 shows a fairly good fit to the actual AIDS incidence by the Bayesian model. This is also seen in Figure 1 which is the plot of the probability of AIDS in each interval. Unlike Blackstone, Naftel and Turner (1986) our hazards are not overlapping. Thus, the Bayesian curve (solid line) is not smooth. The model predicts between 12,663 and 17,743 cases for future time period i = 19, or the latter half of 1987, for a range of the prior parameters. The actual value is approximately 14,000.

### 1.3 Summary

The time varying hazard for the Bayesian model fit reasonably to the empirical data. Prediction is accurate for near future events only. No attempt was made to adjust the model for under-reporting. We are in the process of updating incidence both nationally and locally to test the fit of the model or develop the model further for these events. Note here we modeled the incidence function parameter with the gamma family of distributions. The coherent prior to posterior analysis demonstrated by Bartolucci and Dickey (1977) and by Bartolucci and Singh (1993) held up here as well. Our goal is to further test exponential families of prior distributions with the incidence function. Future studies with the prior beta-binomial family of functions for the discrete probability of AIDS in given time intervals will be examined as well. Gunel (1984) demonstrated the feasibility of this approach in a multinomial model.

The procedure in this paper demonstrates that the incorporation of uncertainty into the infection model is a feasible approach to modeling the incidence of AIDS. As noted above a refinement of the process requires further study.

### 1.4 Acknowledgements

The authors wish to thank Ms. Rita Reid for her time and patience in the typing of this manuscript.

## 1.5 References


Anderson, R.M., Medley, G.F., May, R.M. and Johnson, A.M., A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA J. of Math. Applied in Medicine and Biology*, 3, 229-263, 1986.

Bartolucci, A.A. and Dickey, J.M., Comparative Bayesian and traditional inference for gamma modeled survival data. *Biometrics*, 33(2), 343-354, 1977.

Bartolucci, A.A., Katholi, C.R. and Birch, R., Interim analysis of failure time data - a Bayesian approach. *Environmetrics*, 3(4), 465-477, 1992.

Bartolucci, A.A. and Singh, K.P., A Bayesian approach to bioequivalence in environmental studies. International Congress on Modelling and Simulation. *Proceedings*, 4, 1821-1830, 1993.

Birch R. and Bartolucci, A.A., Determination of the hyperparameters of a prior probability model in survival analysis. *Computers Programs in Biomedicine*, 17, 89-94, 1983.

Blackstone, E.H., Naftel, D.C. and Turner, M.E., The decomposition of time varying hazard into phases, each incorporating a separate stream of concomitant information. *J. Amer. Stat. Assoc.*.

Brookmeyer, R. and Gail, M.H., A method for obtaining short term projections and lower bounds on the size of the AIDS epidemic. *J. Amer. Stat. Assoc.*, 83, 301-308, 1988.

Gunel, E., A Bayesian analysis of the multinomial model for a dichotomous response with non-respondents. *Communications in Statistics-Theory and Methods*, 13(6), 737-751, 1984.

Rosenberg, P.S., Gail, M.H. and Pee, D., Mean square error of estimates of HIV prevalence and short-term AIDS projections derived by backcalculation. *Statistics in Medicine*, 10, 1167-1180, 1991.

Rosenberg, P.S., Bigger, R.J., Goedert, J.J. and Gail, M.H., Backcalculation of the number with human immunodeficiency virus infection in the United States. *Amer. J. of Epid.*, 133(3), 276-285, 1991.

Taylor, J.M.G., Models for the HIV infection and AIDS epidemic in the United States. *Stat. in Med.*, 8, 45-58, 1989.


## 1.6 Tables

| Period $l$ | Time | $\gamma_l$ | $\tau_l$ | Prior | Posterior |
|---|---|---|---|---|---|
| 1 | 1 - 6 | 3.0 | 0.5 | 0.179 | 0.001 |
| 2 | 7 - 9 | 3.0 | 0.5 | 0.308 | 0.007 |
| 3 | 10 - 13 | 3.01 | 1.00 | 0.833 | 0.056 |
| 4 | 14 - 16 | 3.3 | 1.60 | 1.124 | 0.146 |
| 5 | 17 - 18 | 3.1 | 2.1 | 1.238 | 0.304 |

**Table 1**
**Gamma Prior and Posterior Hazards**
**Defined by Hyperparameters $\gamma_l$ and $\tau_l$**

| i = 1 | Interval | Actual Incidence | Bayesian Model |
|---|---|---|---|
| 1 - 3 | 1978(B) - 1979(B) | 13 | 15 |
| 4 | 1980(A) | 20 | 20 |
| 5 | 1980(B) | 34 | 47 |
| 6 | 1981(A) | 89 | 88 |
| 7 | 1981(B) | 181 | 279 |
| 8 | 1982(A) | 368 | 381 |
| 9 | 1982(B) | 656 | 485 |
| 10 | 1983(A) | 1240 | 1196 |
| 11 | 1983(B) | 1616 | 2332 |
| 12 | 1984(A) | 2526 | 2660 |
| 13 | 1984(B) | 3339 | 2968 |
| 14 | 1985(A) | 4717 | 4884 |
| 15 | 1985(B) | 6168 | 6303 |
| 16 | 1986(A) | 7876 | 7426 |
| 17 | 1986(B) | 9456 | 9599 |
| 18 | 1987(A) | 11543 | 11529 |

**Table 2**
**Number of AIDS Cases in Each Six-Month Interval**
**A = January - June B = July - December**

1.6 Figure

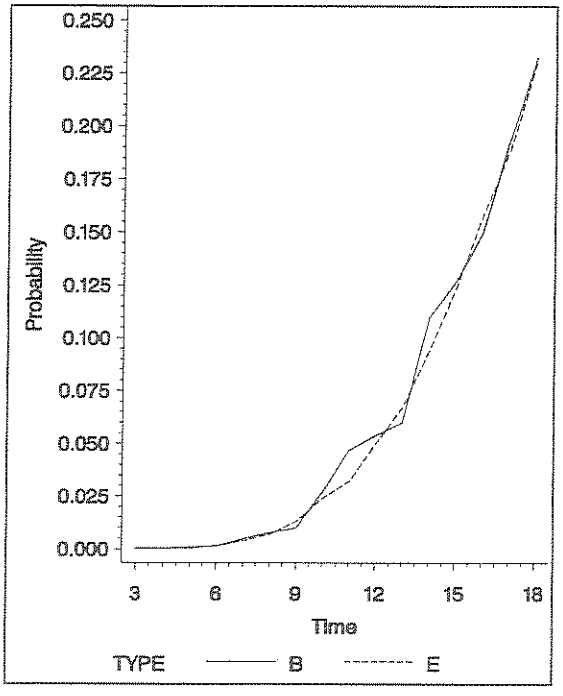


Figure 1
Plot of Empirical (E) Probability of AIDS
vs. Bayesian (B) Solution