Comment

Ron Brookmeyer

I would like to congratulate the authors on an excellent and comprehensive discussion of backcalculation of HIV-infection rates. Bacchetti, Segal and Jewell (BSJ) develop a framework for backcalculation that incorporates several new attractive features, and they discuss important sources of uncertainty.

The earliest data on the incubation period of HIV infection came from a 1986 study of tranfusion-associated AIDS cases. These data indicated that the incubation period was long and variable. This was alarming because it suggested that the number of diagnosed AIDS cases must be only a fraction of the numbers of individuals who were HIV infected. This was the kernel of the idea behind the backcalculation methodology. Brookmeyer and Gail (1986) introduced and applied the methodology to the U.S. AIDS epidemic based on cases diagnosed through 1985 and concluded that even without accounting "for new infections after 1985 nor very long incubation periods," the cumulative number of AIDS cases would grow by a factor of more than 6 by the end of 1991.

Backcalculation is essentially a deconvolution problem. Using data on the cumulative numbers of AIDS cases $a(t)$, and the incubation distribution F, one tries to glean information about past infection rates $I(s)$ through the convolution equation

$$a(t) = \int_{-\infty}^{t} I(s)\, F(t - s \mid s) ds. \tag{1}$$

The solution of (1) has a long and rich statistical history (O'Sullivan, 1986; Wahba, 1990). Equation (1) arises in a wide range of applications, including geology, meterology, engineering and biomedical applications. An important issue that arises in solving (1) concerns how much structure to impose on $I(s)$. Additional problems associated with applying Equation (1) to the AIDS epidemic are imprecise knowledge of the incubation distribution F and systematic errors in AIDS-incidence data.

1. NONSTATIONARY INCUBATION DISTRIBUTION

Backcalculation analyses must account for changes over calendar time in the incubation distribution (Gail, Rosenberg and Goedert, 1990). For example, failure to account for lengthening incubation periods, perhaps because of new treatments, can lead to underestimation of the numbers of infected individuals. One approach to account for nonstationarity effects is to use a completely external estimate of the nonstationary incubation distribution $F(t|s)$ [the treatment model, Brookmeyer and Liao (1990b); Brookmeyer (1991)].

Alternatively, BSJ propose a methodology for estimating nonstationarity effects in the incubation distribution using AIDS-incidence data and backcalculation methods (Section 3.3). However, it seems to be asking a lot of AIDS-incidence data to provide information about both infection rates and changes in the incubation distribution. A falloff in the growth of AIDS cases could either be explained by the scenario of declining infection rates or a scenario of lengthening incubation periods. Intuitively, I would not expect that AIDS-incidence data alone could distinguish between these two scenarios.

The model proposed by BSJ appears to be "nearly nonidentifiable" in the following sense. Given any $\underset{\sim}{\theta}$ and $\underset{\sim}{\beta}$, there exists another $\underset{\sim}{\theta}^*$ and $\underset{\sim}{\beta}^*$ that produce the same likelihood, that is, $L_m(\underset{\sim}{\theta}, \underset{\sim}{\beta}) = L_m(\underset{\sim}{\theta}^*, \underset{\sim}{\beta}^*)$. To see this, assume $\beta_j = 0$ for $j < T$ (BSJ used T = January 1986). Now let $\theta_j^* = \theta_j$ for $j < T$, and choose θ_j^* for $j \geq T$ to be any arbitrary values you like. If we set $\beta_j^* = 0$ for $j < T$, and set

$$\beta_j^* = \beta_j + \log\left\{\left(\sum_{i=0}^{j}\theta_i D_{ij}\right)\Big/\left(\sum_{i=0}^{j}\theta_i^* D_{ij}\right)\right\} \tag{2}$$

for $j \geq T$, then $L_m(\underset{\sim}{\theta}, \underset{\sim}{\beta}) = L_m(\underset{\sim}{\theta}^*, \underset{\sim}{\beta}^*)$. Equation (2) was derived by setting $E(Y_j \mid \underset{\sim}{\theta}, \underset{\sim}{\beta}) = E(Y_j \mid \underset{\sim}{\theta}^*, \underset{\sim}{\beta}^*)$. Equation (2) shows that if one wanted to keep $E(Y_j)$ fixed, lower infection rates ($\theta_j^* < \theta_j$) can be compensated by additional shortening of the incubation period ($\beta_j^* > \beta_j$). On the other hand, higher infection rates ($\theta_j^* > \theta_j$) can be compensated by additional lengthening of the incubation periods ($\beta_j^* < \beta_j$) in order to keep $E(Y_j)$ fixed. This result on nonidentifiability [Equation (2)] holds for the general model proposed by BSJ that incorporates seasonal effects (S_j) and reporting delays (R_j) (setting $S_j^* = S_j$ and $R_j^* = R_j$), as well as for a simple model that assumes no seasonal effects ($S_j = 0$) and with the data completely reported ($R_j = 1$).

BSJ jointly estimate infection rates and nonstationarity effects for the U.S. AIDS epidemic. In this case, simultaneous estimation of $\underset{\sim}{\beta}$ and $\underset{\sim}{\theta}$ appears to be possible solely because of the additional smoothness

Ron Brookmeyer is a Professor in the Department of Biostatistics, School of Hygiene and Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205.