Bayesian Evaluation of CD4 Count as a Surrogate Endpoint in Patients with Advanced HIV Infection

Bradley P. Carlin and Anne I. Goldman
University of Minnesota, USA

SUMMARY

We examine CD4 lymphocyte responses and survival patterns in 467 persons with advanced HIV-1 infection who were randomly assigned to treatment with one of two antiretroviral drugs, didanosine (ddI) and zalcitabine (ddC). Analyses of CD4 lymphocyte count over time are performed using longitudinal random effects models which allow for the possibility of a drug-induced CD4 increase. Probit regression is also used to directly estimate the probability of such an increase in each treatment group. Finally, we assessed the prognostic value of CD4 response and drug group as predictors of survival time. Throughout, Markov chain Monte Carlo methods enable computation of the relevant posterior summaries. We find that while patients given ddI had a substantially greater chance of a CD4 response after two months than those given ddC, only CD4-responding ddC patients experienced improved survival. Our results thus provide compelling evidence against the common practice of using CD4 count as a surrogate marker for survival in persons with advanced HIV disease.

Keywords: AIDS; Longitudinal models; Markov chain Monte Carlo; Probit regression; Survival analysis.

1. INTRODUCTION

Large-scale clinical trials often require extended periods of follow-up to evaluate the clinical efficacy of a new treatment, as measured for example by survival time. The resulting time lags in reporting the results of the trial to the clinical and patient communities are especially problematic in AIDS research, where the life-threatening nature of the disease
heightens the need for rapid dissemination of knowledge. A common approach for dealing with this problem is to select an easily-measured biological marker, known to be predictive of the clinical outcome, as a surrogate endpoint. For example, the number of CD4 lymphocytes per cubic millimeter of drawn blood has been used extensively as a surrogate marker for progression to AIDS and death in drug efficacy studies of HIV-infected persons. Indeed, an increase in the average CD4 count of patients receiving one of the drugs in our study (didanosine) in a clinical trial was the principal argument used in its successful bid for limited licensing in the United States.

Several investigators, however, have questioned the appropriateness of using CD4 count in this capacity. For example, Lin et al. (1993) showed that CD4 count was not an adequate surrogate for first opportunistic infection in data from the Burroughs Wellcome 02 trial, nor was it adequate for the development of AIDS or death in AIDS Clinical Trials Group (ACTG) Protocol 016 data. Choi et al. (1993) found CD4 count to be an incomplete surrogate marker for AIDS using results from ACTG Protocol 019 comparing progression to AIDS for two dose groups of zidovudine (AZT) and placebo in asymptomatic patients. Most recently, the results of the joint Anglo-French “Concorde” trial (Concorde Coordinating Committee, 1994) showed that while asymptomatic patients who began taking AZT immediately upon randomization did have significantly higher CD4 counts than those who deferred AZT therapy until the onset of AIDS-related complex (ARC) or AIDS itself, the survival patterns in the two groups were virtually identical.

We analyze data from a trial involving 167 persons with advanced HIV infection who were randomly assigned to treatment with the antiretroviral drugs didanosine (ddI) or zalcitabine (ddC). The details of the conduct of the ddI/ddC study are described elsewhere (Abrams et al., 1994); only the main relevant points are given here. The trial enrolled HIV-infected patients with AIDS or two CD4 counts of 300 or less, and who fulfilled specific criteria for
AZT intolerance or failure. CD4 counts were recorded at study entry and again at the 2, 6, 12, and 18 month visits (though some of these observations are missing for many individuals). The study measured several outcome variables; we consider only total survival time.

Our main goal is to analyze the association among CD4 count, survival time, drug group, and AIDS diagnosis at study entry (an indicator of disease progression status). Our population of late-stage patients is ideal for this purpose, since it contains substantial information on both the actual (survival) and surrogate (CD4 count) endpoints. Section 2 discusses our approach for modeling the series of longitudinal CD4 counts, while Section 3 attempts to model a drug-induced response in CD4 count directly using probit regression. Section 4 then compares the survival times of patients in each drug and CD4 response group. Finally, Section 5 summarizes our findings and discusses their implications for patient care.

2. MODELING OF LONGITUDINAL CD4 COUNTS

Suppose we let \( Y_{ij} \) denote the \( j^{th} \) CD4 measurement on the \( i^{th} \) individual in the study, \( j = 1, \ldots, s_i \) and \( i = 1, \ldots, n \). Arranging each individual’s collection of observations in a vector \( Y_i = (Y_{i1}, \ldots, Y_{is_i})^T \), we might attempt to fit the random effects model

\[
Y_i = X_i \alpha + W_i \beta_i + \epsilon_i,
\]

where \( X_i \) is a \( s_i \times p \) design matrix, \( \alpha \) is a \( p \times 1 \) vector of fixed effects, \( W_i \) is a \( s_i \times q \) design matrix (\( q \) typically less than \( p \)), and \( \beta_i \) is a \( q \times 1 \) vector of subject-specific random effects, usually assumed to be normally distributed with mean vector \( \mathbf{0} \) and covariance matrix \( \mathbf{V} \). We assume that given \( \beta_i \), the components of \( \epsilon_i \) are independent and normally distributed with mean 0 and variance \( \sigma^2 \) (marginalizing over \( \beta_i \), the \( Y_i \) components are again correlated, as desired). The \( \beta_i \)'s capture any subject-specific mean effects, and also enable the model to
reflect any extra-normal variability in the data. Models of this type have been very popular for longitudinal data since their appearance in the paper of Laird and Ware (1982).

Since we wish to detect a possible increase in CD4 count two months after baseline, we attempt to fit a model that is linear but with possibly different slopes before and after this time. Thus the subject-specific design matrix $W_i$ for patient $i$ in equation (1) has $j^{th}$ row $w_{ij} = (1, t_{ij}, (t_{ij} - 2)^+)$, where $t_{ij} \in \{0, 2, 6, 12, 18\}$ and $z^+ = \max(z, 0)$. Hence the three columns of $W_i$ correspond to individual-level intercept, slope, and change in slope following the changepoint, respectively. We account for the effect of covariates by including them in the fixed effect design matrix $X_i$. Specifically, we set $X_i = (W_i | d_iW_i | a_iW_i)$, where $d_i$ is a binary variable indicating whether patient $i$ received ddI ($d_i = 1$) or ddC ($d_i = 0$), and $a_i$ is another binary variable telling whether the patient was diagnosed as having AIDS at baseline ($a_i = 1$) or not ($a_i = 0$). Notice that we have $p = 3q = 9$; the two covariates are being allowed to affect any or all of the intercept, slope, and change in slope of the overall population model. The corresponding elements of the $\alpha$ vector then quantify the effect of the covariate on the form of the CD4 curve. In particular, our interest focuses on the $\alpha$ parameters corresponding to drug status, and whether they differ from 0.

Adopting the usual exchangeable normal model for the random effects, we obtain a likelihood of the form

$$
\prod_{i=1}^{n} N_{s_i}(Y_i|X_i\alpha + W_i\beta_i, \sigma^2 I_{s_i}) \prod_{i=1}^{n} N_3(\beta_i|0, V),
$$

where $N_k(\cdot | \mu, \Sigma)$ denotes the $k$-dimensional normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$. To complete our Bayesian model specification, we adopt the prior distributions $N_3(\alpha|c, D)$, $IG(\sigma^2|a, b)$, and $IW(V|\rho R)^{-1}$, where $IG$ and $IW$ denote the inverse gamma and inverse Wishart distributions, respectively.
Figure 1: Exploratory plots of CD4 count, ddI/ddC data

Turning to the observed data in our study, boxplots of the individual CD4 counts for the two drug groups, shown in Figures 1(a) and (b), indicate a high degree of skewness toward high CD4 values. This, combined with the count nature of the data, suggests a square root transformation for each group. As Figures 1(c) and (d) show, this transformation improves matters considerably. The sample medians, shown as white horizontal bars on the boxplots, offer reasonable support for our assumption of a linear decline in square root CD4 after two months. The sample sizes at the five time points, namely (230, 182, 153, 102, 22) and (236, 186, 157, 123, 14) for the ddI and ddC groups, respectively, indicate an increasing degree of
missingness as the study wears on. There are only 1405 total observations, for an average of roughly 3 per study participant.

We chose hyperparameter values that specified vague prior distributions on $\sigma^2$, $\alpha$, and $\mathbf{V}$, and obtained posterior summaries via the Gibbs sampler. In the interest of brevity we omit further details concerning prior specification and algorithm implementation, instead referring the reader to the forthcoming paper by Carlin (1994).

We obtained 95% equal-tail posterior credible sets of (.074, .580) and (-.671, -.074) for $\alpha_5$ (pre-changepoint drug slope) and $\alpha_6$ (post-changepoint drug slope), respectively. This suggests that the CD4 trajectories of persons receiving ddI have slopes that are significantly different from those of the ddC patients, both before and after the two-month changepoint. Plots of the fitted population CD4 trajectories for each of the four possible drug-diagnosis combinations (not shown) suggest that an improvement in square root CD4 count typically occurs only for AIDS-negative patients receiving ddI. However, there is also some indication that the ddC trajectories “catch up” to the corresponding ddI trajectories by the end of the observation period. Moreover, while the trajectories of ddI patients may be somewhat better, the difference is quite small, and probably insignificant clinically.

3. CD4 RESPONSE TO TREATMENT AT TWO MONTHS

The longitudinal changepoint model of the previous section takes advantage of all the observations for each individual, but provides only an indirect way of capturing the chance of a CD4 response in each drug group. As an alternative, we reclassify each patient as “response” or “no response,” depending on whether or not the CD4 count increased from its baseline value. That is, for $i = 1, \ldots, m = 367$, the number of patients who had CD4 measured at both the baseline and two month visits, we define $R_i = 1$ if $Y_{i2} - Y_{i1} \geq 0$, and $R_i = 0$ otherwise. Defining $p_i$ as the probability that patient $i$ responded, we fit the probit
regression model
\[ p_i = \Phi(\gamma_0 + \gamma_1 d_i + \gamma_2 a_i), \]  
(2)

where $$\Phi$$ denotes the cumulative distribution function of a standard normal random variable, the $$\gamma$$'s are unknown coefficients, and the covariates $$d_i$$ and $$a_i$$ describe treatment group and baseline AIDS diagnosis as in Section 2. We used probit regression in place of the similar (but more common) logistic regression approach since the former is more easily fit using the Gibbs sampler (Carlin and Polson, 1992; Albert and Chib, 1993).

Under a flat prior on $$(\gamma_0, \gamma_1, \gamma_2)'$$, the posterior distributions for each of the $$\gamma$$ coefficients were used to derive point and interval estimates of the corresponding covariate effects. In addition, given particular values of the covariates, the posterior distributions were used with equation (2) to estimate the probability of a CD4 response in subgroups of patients by treatment and baseline prognosis. The point and 95% interval estimates for the regression coefficients obtained using the Gibbs sampler are shown in Table 1. Of the three intervals, only the one for prior AIDS diagnosis excludes 0, though the one for treatment group nearly does. The signs on the point estimates of these coefficients indicate that patients without a prior AIDS diagnosis and, to a lesser extent, those in the ddl group were more likely to experience a CD4 response, in agreement with our Section 2 results.

<table>
<thead>
<tr>
<th></th>
<th>point estimate</th>
<th>95% confidence limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$$\gamma_0$$ (intercept)</td>
<td>.120</td>
<td>-.135 -.378</td>
</tr>
<tr>
<td>$$\gamma_1$$ (treatment)</td>
<td>.226</td>
<td>-.040 .485</td>
</tr>
<tr>
<td>$$\gamma_2$$ (AIDS diagnosis)</td>
<td>-.339</td>
<td>-.610 -.068</td>
</tr>
</tbody>
</table>

Table 1: Point and interval estimates, CD4 response model coefficients

Converting these results for the regression coefficients to the probability-of-response scale using equation (2), we may again compare results for a typical patient in each drug-diagnosis
group. Table 2 shows that the posterior median probability of a response is larger by almost .09 for persons taking ddI, regardless of health status. If we compare across treatment groups instead of within them, we see that the patient without a baseline AIDS diagnosis has roughly a .135 larger chance of responding than the sick patient, regardless of drug status. Notice that for even the best response group considered (baseline AIDS-free patients taking ddI), there is a substantial estimated probability of not experiencing a CD4 response, confirming the rather weak evidence in favor of its existence under the full random effects model (1).

| AIDS diagnosis at baseline | ddI | ddC | No AIDS diagnosis at baseline | .637 | .550 |

Table 2: Point estimates, probability of response given treatment and prognosis

4. SURVIVAL ANALYSIS

In order to evaluate the clinical consequences of two-month changes in the CD4 lymphocyte count associated with the study drugs, we performed a parametric survival analysis. As in Section 3, the $m = 367$ patients with baseline and two-month CD4 counts were divided into responders and non-responders. We wish to fit a proportional hazards model, wherein the hazard function $h$ for a person with covariate values $\mathbf{z}$ and survival or censoring time $t$ takes the form $h(t|\mathbf{z}, \boldsymbol{\beta}) = h_0(t) \exp(\mathbf{z}'\boldsymbol{\beta})$. In our case, we define four covariates as follows: $z_0 = 1$ for all patients; $z_1 = 1$ for ddI patients with a CD4 response, and 0 otherwise; $z_2 = 1$ for ddC patients without a CD4 response, and 0 otherwise; and $z_3 = 1$ for ddC patients with a CD4 response, and 0 otherwise. Writing $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ and following Cox and Oakes (1984, Sec 6.1), we obtain the loglikelihood

$$\log L(\boldsymbol{\beta}) = \sum_{i \in \tilde{t}} \log h(t_i|\mathbf{z}_i, \boldsymbol{\beta}) + \sum_{i=1}^m \log S(t_i|\mathbf{z}_i, \boldsymbol{\beta}),$$

(3)
where $S$ denotes the survival function, $\mathcal{U}$ the collection of uncensored failure times, and $z_i = (z_{0i}, z_{1i}, z_{2i}, z_{3i})'$. Our parametrization uses nonresponding ddI patients as a reference group; $\beta_1$, $\beta_2$ and $\beta_3$ capture the effect of being in one of the other 3 drug-response groups.

We followed Dellaportas and Smith (1993) by beginning with a Weibull model for the baseline hazard, $h_0(t) = pt^{\rho-1}$, but replaced their rejection sampling algorithm based on the concavity of the loglikelihood (3) with the easier-to-program Metropolis subchain approach (see e.g. Müller, 1994). Our initial concern that the Weibull model might not be rich enough evaporated when we discovered extremely high posterior correlation between $\rho$ and $\beta_0$. This suggests the baseline hazard may reasonably be thought of as exponential in our very ill population, and hence we fixed $\rho = 1$ in all subsequent calculations.

The resulting estimated marginal posterior distributions for the $\beta$’s were fairly symmetric, and those for $\beta_1$ and $\beta_2$ were centered near 0. However, the 95% equal-tail posterior credible set for $\beta_3$, $(-1.10, 0.02)$, suggests predominantly negative values. To ease the interpretation of this finding, we transform the posterior $\beta$ samples into corresponding ones from the survival function using the relation $S(t|z, \beta) = \exp\{-t \exp(z'\beta)\}$. We do this for each drug-response group over a grid of 9 equally-spaced $t$ values from 0 to 800 days. Figure 2(a) gives a smoothed plot of the medians of the resulting samples, thus providing estimated posterior survival functions for the four groups. As expected, the group of ddC responders stands out, with substantially improved mortality. A more dramatic impression of this difference is conveyed by the estimated posterior distributions of median survival time, $\theta(z) = (\log 2) \exp(-z'\beta)$, shown in Figure 2(b). While a CD4 response translates into improved survival in both drug groups, the improvement is clinically significant only for ddC recipients.

5. DISCUSSION

Analysis of our CD4 data using both a longitudinal changepoint model and a simpler
probit regression model suggests that ddC is less successful than ddI in producing a CD4 boost in patients with advanced HIV infection. However, this superior CD4 performance does not seem to translate into improved survival for ddI patients; in fact, it is the CD4-responding ddC patients who attain significant gains in survival. While CD4 count has undeniable value as a prognostic indicator for clinicians, it appears that modifying its measured level using antiretroviral drugs does not have a predictable effect on survival. Hence our study, like Concorde and others in the recent literature, calls the value of CD4 as a surrogate endpoint
into question. It suggests rethinking the practice of licensing drugs primarily on the basis of a demonstrated increase in CD4 count. Moreover, given the unpleasant side effects associated with these drugs, review of their usage with end-stage patients seems warranted.

6. ACKNOWLEDGEMENTS

Partial support for this work was provided by CPCRA Contract NO1-A1-05073 from the Division of AIDS, National Institute of Allergy and Infectious Diseases (NIAID). In addition, the work of Carlin was supported in part by NIAID FIRST Award 1-R29-AI33466.

REFERENCES


