

# Complexity of large observational studies: missing data, competing risks, association and causality

Daniela Zugna<sup>1</sup>, Bianca De Stavola, Ronald Geskus, and Kholoud Porter  
Vita-Salute San Raffaele University, Milano, e-mail: zugna.daniela@hsr.it

**Abstract:** The risk of treatment failure following initiation of HAART (highly active antiretroviral therapy) close to seroconversion in HIV infected persons, is of interest to clinicians considering early treatment of seroconverters. Indeed it is not known whether this risk is different to that seen among patients starting treatment during chronic infection. Evaluation is not trivial since treatment cessation/interruption acts as a competing risk, further assignment to earlier or later treatment is not random, and biomarkers measurements are usually available at irregular time intervals, at varying time points and in unequal numbers for different study participants. By using a large dataset from the CASCADE Collaboration of HIV infected persons with well-estimated dates of seroconversion, we show an example of this framework and we propose a possible solution to these problems.

**Keywords:** observational study, inverse probability of treatment weights, bivariate linear mixed effects model, competing risks

## 1. Introduction

The field of HIV research has been profoundly altered since the introduction of HAART (highly active antiretroviral therapy) in 1995. This treatment led to a substantial reduction in mortality and disease progression to AIDS by suppressing viral load and increasing CD4 cell count. Yet, a debate regarding the optimal timing of initiation of HAART in chronically HIV-infected individuals is in progress. The main arguments for starting HAART early is to reduce the risk of opportunistic infections, with the consequent improvement in quality of life, and to preserve HIV specific cellular response. In contrast, while the short- and medium-term side-effects of HAART are known, the long-term toxicity is not, and therefore the benefits of delayed treatment would include instead avoiding the side-effects of lifelong treatment with antiretroviral drugs and minimizing the development of viral drug resistance. Hence the aim of this study was to evaluate if the risk of treatment failure in patients who start HAART within 12 months since seroconversion (early treated) is different to that seen among patients starting treatment during chronic infection (late treated). Since one goal of HAART is to achieve undetectable plasma HIV RNA, we defined the “timing of treatment failure” as the date of first measurement in HIV RNA of  $> 1000$  copies/ml after previous viral suppression to  $< 500$  copies/ml or HIV RNA of  $> 500$  copies/ml after 6 months following HAART initiation.

Such an evaluation is not trivial due to the nature of HIV data: the survival component involves several sources of bias, non-random allocation of treatment, multiple time-scales and informative censoring, while the longitudinal component includes important biomarkers, which are irregularly and poorly measured.

## 2. Modeling and estimation

### 2.1 Survival model

Since the follow-up was artificially censored at the first major modification or at the interruption of the therapy, defined as date which HAART was modified or interrupted for more than one week respectively, we considered them as competing risks (Kalbfleisch and Prentice, 2002; Putter et al., 2007). In this framework, a competing risk is an event whose occurrence alters the probability of occurrence of other events, and appropriate methodology has to be used. By using the Fine and Gray model (Fine and Gray, 1999), we modelled the complementary log-log transformation of the cumulative incidence function for the event of interest,  $L = 1$ , defined as:

$$I_1(t; \mathbf{z}) = P(T < t, L = 1 | \mathbf{z})$$

It is interpreted as the marginal cumulative probability of a particular event occurring as first one conditional on covariates  $\mathbf{z}$ . Specifically we modelled not the usual hazard function, but the subdistribution hazard function for failure type  $l$ , given by:

$$\lambda_l(t; \mathbf{z}) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t, L = l | T > t \cup (T < t, L \neq l))}{\Delta t}$$

Hence the estimated subdistribution hazard ratio of early HAART on late HAART can be interpreted as the instantaneous effect of starting therapy within a year since seroconversion on starting later conditional on having survived only the event of interest. Through prediction procedures it is possible to evaluate the direct effect of each covariate directly on the cumulative incidence curves.

Since the analyses was carried out on time since HAART initiation and treatment was not assigned at random, it was needed to remove the association between the factors that influenced treatment allocation and treatment effect, by inverse-probability-of-treatment weighed (IPTW) estimators (Hernan et al., 2001). The inverse probability of treatment weight at time  $t$  is defined as:

$$SW(t) = \prod_{k=0}^t \frac{f(A(k) | \bar{A}(k-1), V)}{f(A(k) | \bar{A}(k-1), \bar{L}(k))}$$

where  $A(t)$  is the treatment at time  $t$ ,  $\bar{A}(t)$  is the treatment history until time  $t$ ,  $\bar{L}(t)$  is the prognostic factors history until time  $t$ , and  $V = L(0)$ . Hence this approach consists of estimating each subject's probability of having their own treatment history at time  $t$  conditional on past treatment and prognostic factors history. Specifically we considered sex, HIV risk group, age at seroconversion, calendar year of HAART initiation, and CD4 and HIV RNA patterns as relevant prognostic factors. In order to obtain well-behaved weights, the truncation of inverse probability weights was applied, by ranking them below the first centile and above the 99th centile.

Now the subdistribution hazards model was weighed with IPTW in order to estimate the causal effect of exposure on competing events. Further variables of interest, such as sex, HIV risk group, age at seroconversion, calendar year of HAART initiation, type of regimen, and CD4 and HIV RNA count at cART initiation, were included in the model to evaluate their association with the probability of failing the treatment, changing treatment regimen and interrupting the therapy.

## 2.2 Longitudinal model

Since CD4 and viral load measurements are available at irregular time intervals and in unequal numbers for different study participants, their values were predicted using the available information on their history from seroconversion up to HAART initiation. In order to normalize the biomarkers distribution, the square root of CD4 count and the log transformation of HIV RNA count were fitted. The biomarkers pattern was modelled by using a bivariate linear mixed effects model (Diggle et al., 2002), given by:

$$\begin{cases} y'_{ij} = u'_{ij} + \epsilon'_{ij} \\ y''_{ij} = u''_{ij} + \epsilon''_{ij} \end{cases}$$

where

$$\begin{aligned} u'_{ij} &= \alpha'_1 + \alpha'_2 I(exposure) + \alpha'_3 t_{ij} + \theta_{1i} + \theta_{2i} t_{ij} \\ u''_{ij} &= \alpha''_1 + \alpha''_2 I(exposure) + \alpha''_3 t_{ij} + \theta_{3i} + \theta_{4i} t_{ij} \end{aligned}$$

$y'_{ij}$  and  $y''_{ij}$  are the square root of cd4 cell count and the log transformation of viral load at time  $t_{ij}$  for the  $i$ th subject, respectively. Because viral load cannot be measured below a certain limit, depending on the tool used to measure it, we considered the left truncation of the biomarker in modelling it. The parameters  $\alpha'$  and  $\alpha''$  are  $3 \times 1$  vectors of unknown fixed effects,  $\theta_i$  is a  $4 \times 1$  vector of unobservable random effects, and  $\epsilon'_i$  and  $\epsilon''_i$  are within-individual residuals vectors. The random effects  $\theta_i$  are assumed to be normally distributed with mean 0 and  $4 \times 4$  variance-covariance matrix  $\Sigma$ . The  $\theta_i$  are distributed independent of each other and of the within-subjects residuals  $\epsilon'_{ij}$  and  $\epsilon''_{ij}$ . The  $\epsilon'_i$  are assumed to be independent and normally distributed with mean 0 and  $n'_i \times n'_i$  variance-covariance matrix  $\sigma'^2_{\epsilon} I_{n'_i}$ , where  $n'_i$  is the number of measurements for subject  $i$  and  $I_{n'_i}$  denotes the  $n'_i \times n'_i$  identity matrix. Likewise for  $\epsilon''_i$ . We modelled the dependence between viral load and CD4 cell count by the covariance matrix of the individual effects  $\theta_i$ ,  $\Sigma$ . The parameters  $\alpha'_2$  and  $\alpha''_2$  represent the difference in biomarkers' trend between exposed and unexposed. By a Bayesian approach, proper priors were selected but with hyperparameter values chosen so that the priors had a minimal impact on the posterior estimates. Specifically multivariate normal priors were adopted for the main effects vector, an inverse gamma priors for the error variance, and an inverse Wishart for the covariance-variance matrix, all having very low precision (Hastings, 1970).

## 3. Application: AIDS data

We used data collected in 2007 from the CASCADE Collaboration (Concerted Action on Seroconversion to AIDS or Death in Europe) of HIV infected persons with well-estimated dates of seroconversion, CD4 and HIV RNA measurements, all anti-retroviral drugs prescribed, vital status and date of last clinical assessment. The final analysis was based on 1337 patients, 984 (73.60%) late treated versus 353 (26.40%) early treated. Late treated patients were more likely to suffer from treatment failure (20.1% versus 16.4%) than early treated patients, while they were less likely to change (26.42% versus 30.59%) and to interrupt the therapy (15.45% versus 33.71%).

The estimated mean intercepts at seroconversion for square root CD4 count equalled to 24.09 (95% IC 23.69 to 24.48) and 22.62 (95% IC 21.41 to 23.83) cells/ $\mu$ L for late and

early treated patients, respectively. The corresponding values for log HIV RNA count were 4.11 (95% IC 4.03 to 4.20) and 5.01 (95% IC 4.87 to 5.39)  $\log_{10}$  copies/mL. The estimated mean CD4 count slope was -0.128 (95% IC -0.160 to -0.010) square root cells/ $\mu$ L per month, while the estimated mean viral load slope was 0.011 (95% IC -0.004 to 0.027)  $\log_{10}$  copies/mL per month. As expected, the average CD4 count decreased before cART initiation, while HIV RNA count increased in both groups.

After accounting for the other two competing events, calculating IPTW, and controlling for possible confounders, the adjusted subdistribution hazard for treatment failure and treatment change turned out similar for early and late treated individuals respectively (sub-HR=0.889,  $p=0.490$ ; sub-HR=1.088,  $p=0.540$ ), and the adjusted subdistribution hazard of treatment interruption was much higher for early compared to late treated individuals (sub-HR=1.890,  $p= <0.001$ ). We report the results of subdistribution hazard ratios for treatment failure in table 1.

**Table 1:** *Adjusted sub-hazard ratio for treatment failure*

|                                  | sub-HR | Conf. Int (95%) | p-value |
|----------------------------------|--------|-----------------|---------|
| Exposed/Unexposed                | 0.889  | (0.637,1.241)   | 0.490   |
| Female/Male                      | 1.071  | (0.616,1.241)   | 0.809   |
| MSW/MSM                          | 0.911  | (0.577,1.439)   | 0.690   |
| other/MSM                        | 0.419  | (0.135,1.296)   | 0.131   |
| 25-34/15-24                      | 0.577  | (0.424,0.785)   | <0.001  |
| 35-44/15-24                      | 0.738  | (0.517,1.052)   | 0.093   |
| 45+/15-24                        | 0.280  | (0.156,0.502)   | <0.001  |
| PI/NNRTI                         | 0.820  | (0.610,1.102)   | 0.188   |
| 3N/NNRTI                         | 1.217  | (0.719,2.058)   | 0.465   |
| 3 CLASS/NNRTI                    | 0.419  | (0.096,1.826)   | 0.247   |
| HAART year                       | 0.825  | (0.771,0.882)   | <0.001  |
| cd4 baseline (square root scale) | 0.990  | (0.965,1.016)   | 0.455   |
| rna baseline (log scale)         | 1.238  | (1.053,1.457)   | 0.010   |

We found evidence to suggest that the youngest age group had the highest risk of treatment failure and that our oldest age group, 45 years or more at seroconversion, was less likely to fail compared to 16-24 year olds given the weaker immune system of the former. The risk of treatment failure decreased over calendar time, even after controlling for changes which took place over that period in HAART classes. Finally, the risk of failure was higher at high HIV RNA at baseline.

## 4. Discussion

Large observational studies are characterized by several sources of bias, due to longitudinal and survival nature of data and study design. Focusing on the clinical question of interest on the optimal time when to initiate the therapy, we propose a method which models time-dependent variables in order to overcome the problem of missing data over time and accounts for individual heterogeneity, calculates the weights in order to create a pseudo-population where the treatment is supposed to be allocated randomly and independently of individual status, accounts for competing events in order to use the information given by censoring, and quantifies the association between the risk of failure and the variables of interest, by an appropriate regression model.

## References

- Fine, J. and Gray, R. (1999) A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, 94, 496-509
- Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002) *Analysis of Longitudinal Data*, Oxford University Press.
- Hastings, W. (1970) Monte carlo sampling methods using markov chain and their application, *Biometrika*, 57, 549-603
- Hernan, M., Brumback, B., Robins, J.M., (2001) Marginal Structural Models to Estimate the Join Causal Effect of Nonrandomized Treatments, *Journal of the American Statistical Association*, 96, 440-448
- Kalbfleisch J.D., Prentice R.L. (2002) *The Statistical Analysis of Failure Time Data* , Wiley Series in Probability and Statistics.
- Putter, H., Fiocco, M., and Geskus, R. (2007) Tutorial in biostatistics: Competing risks and multi-state models, *Statistics in medicine*, 26, 2389-2430