CLUSTERING OF HCV VIRAL LOAD PROFILES IN UNTREATED PATIENTS

Jian Huang [1], Kathleen O'Sullivan[1], John Levis[2], Orla Crosbie[3], Elizabeth Kenny-Walsh[3] and Liam J. Fanning[2]

[1]Statistical Consultancy Unit, University College Cork, Ireland
[2]Molecular Virology Diagnostic and Research Laboratory, Department of Medicine, Cork University Hospital and University College Cork, Ireland.
[3]Department of Gastroenterology and Hepatology, Cork University Hospital, Ireland

The viral load of the hepatitis C virus (HCV) is clinically important as a predictor of response to interferon therapy. A wide dynamic range of viraemia can occur within an untreated patient with chronic hepatitis C [1]. There is currently a paucity of information regarding complex biological parameters that impact on the magnitude of HCV viral load at any given time and how viral load changes over time within an untreated patient population. Using data collected from untreated patients with chronic hepatitis C, we adapt a model based clustering method described by Luan and Li [2] to investigate whether there are different viral load progression patterns.

Considering viral load as a smooth function of time from infection, we treat the observed viral load of each patient as a sample taken from an underlying smooth function. By approximating the underlying smooth function as a linear combination of base functions and treating the coefficients as random effects, we propose a random effects model to fit viral load data. The vector of random coefficients of each patient is assumed from a mixture of the C (multi-dimension) normal distributions with the different means and same covariance matrix. The estimated mean vectors determine cluster profiles which may represent different patterns of viral progression over time. Given C, the EM algorithm is employed to estimate the normal distribution parameters and random effects and unknown cluster memberships of patients. C is selected by the Bayesian Information Criteria. The method has flexibility in the sense that the observation times are not required to be equally spaced for patients, and furthermore, these times can vary from one patient to another.

Data used to demonstrate the proposed method consisted of 85 individuals whose sole risk factor for acquisition of hepatitis C was iatrogenic infection through receipt of HCV genotype 1 subtype b (HCV 1) contaminated Anti-D immunoglobulin, each contributing between 2 to 10 years of measurements. The clustering results indicate that different temporal patterns of viral load exist. These patterns are discussed in light of the recently developed hypothesis of replicative homeostasis and shifts in viral load dynamics [3].

References

[1] Fanning, L., Kenny-Walsh, E., Levis, J., Choudhury, K.R., Cannon, B., Sheehan, M., Whelton, M. and Shanahan, F. (2000). Natural fluctuations of hepatitis C viral load in a homogeneous patient population: a prospective study. Hepatology 31, 225-9.
[2] Luan, Y., Li, H., 2003. Clustering of time-course gene expression data using a mixed-effects model with B-splines. Bioinformatics 19, 474-482.
[3] Sallie R. (2005) Replicative homeostasis: a fundamental mechanism mediating selective viral replication and escape mutation. Virol J. 11, 2:10