Modelling the HIV and AIDS Epidemic with Application to Malawi

Daisy Salifu

A thesis submitted in fulfillment for the Degree of Doctor of Philosophy in Applied Statistics in the Jomo Kenyatta University of Agriculture and Technology

2010

ABSTRACT

Various methods have been developed for modelling the human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS) epidemic to understand the size of the epidemic and the rate of development of new infections overtime. The major objective of this study was to investigate possible epidemic models that could describe and derive trend estimates of the HIV and AIDS epidemic with specific reference to Malawi.

We focussed on two methods namely back-calculation methods and deterministic models.

Two forms of back-calculation methods were considered; parametric and non-parametric back-calculation implemented using the EM-algorithm. In the parametric case, splines and non-linear models were assumed for the infection curve. Estimates obtained using linear, quadratic and cubic splines differed and depended on the number of knots. There were two non-linear models fitted to the AIDS data namely; quadratic exponential and linear logistic. Overall, the cubic spline without any knot fitted the data better than the rest of the models. The back-calculation for the cubic spline suggests HIV incidence peaked in 1990 and then declined by less than 5% each year and peaked again in 1996, but the peak is smaller than that of 1990. Estimates from the non-parametric back-calculation were obtained using the EM-algorithm. The modelled HIV incidence was lower for the nonparametric case as compared to the parametric case. The estimated HIV incidence suggests a peak in 1991 and thereafter declined gradually. The non-parametric bootstrap confidence interval revealed that estimates of recent HIV incidence are poorly determined (high standard error) so that any extrapolations into the future can only be hypothetical.

Sensitivity of the parametric and non-parametric back-calculation was studied by varying the median incubation period of the Weibull distribution. The higher the incubation period the

greater the HIV incidence estimates implying that HIV incidence would be overestimated if the incubation period is overestimated and underestimated if the incubation period is underestimated. This result has far-reaching implications in the modelling of HIV and AIDS epidemic in sub-Saharan Africa, where to date no empirical studies have been conducted to establish the incubation period distribution of AIDS.

Convergence rate of the EM-algorithm was slow at higher median incubation period and fast at low median incubation period thus the greater the information contained in the complete data, the slower the convergence rate.With respect to the deterministic model, the simulations suggests that initially there is an exponential growth in the HIV incidence curve reaching the peak at about 15 years (thus 1996) since the outbreak before decreasing sharply. The model projections seem to be valid up to 25 years since the start of epidemic as beyond 25 years the estimates do not change.

The uncertainty surrounding the incubation period distribution of AIDS, the infection curve and data sources suggests the need to explore methods that allow for additional information known about the epidemic to be incorporated into the back-calculation model.

Bayesian approaches might be useful in this regard.