Statistical Methods for Modeling HIV/AIDS in India

Corresponding Author:
Dr. Brijesh Sathian,
Assistant Professor & Managing and Chief Editor NJE, Community Medicine, Manipal College of Medical Sciences, Department of Community Medicine, Manipal College of Medical Sciences, 155 - Nepal

Submitting Author:
Dr. Brijesh Sathian,
Assistant Professor & Managing and Chief Editor NJE, Community Medicine, Manipal College of Medical Sciences, Department of Community Medicine, Manipal College of Medical Sciences, 155 - Nepal

Article ID: WMC003336
Article Type: Review articles
Submitted on: 05-May-2012, 06:08:59 PM GMT  Published on: 07-May-2012, 04:24:51 PM GMT
Article URL: http://www.webmedcentral.com/article_view/3336
Subject Categories: BIOSTATISTICS
Keywords: Statistical Modeling, HIV/AIDS, Curve fitting, India

How to cite the article:

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Source(s) of Funding:
Not Applicable

Competing Interests:
No competing interests
Statistical Methods for Modeling HIV/AIDS in India

Author(s): Sathian B, Sreedharan J

Abstract

Deterministic, Stochastic, Statistical and State space models are the statistical models for forecasting HIV/AIDS data. There are also uncertainties associated with these approaches. In addition to this the recent advance in this field used is curve fitting models. Sathian B and Sreedharan J used this method for forecasting several infectious and non infectious diseases. It gives more accurate estimates compared to the other models.

Introduction

Adult HIV prevalence in India declined from 0.41% in 2000 to 0.31% in 2009. The 2008-09 India HIV estimates developed by NACO with support from National Institute of Medical Sciences, National Institute of Health and Family Welfare, UNAIDS and WHO utilised improved methodology and updated epidemiological data from the latest rounds of HIV Sentinel Surveillance and other information on High Risk Groups for more accurate understanding of the Indian epidemic. It is estimated that India had approximately 1.2 lakh new HIV infections in 2009, as against 2.7 lakh in 2000[1]. Statistical methods have expanded spatially in recent years to address large scale worldwide health issues. These methods have a prominent role in the study of the HIV/AIDS epidemic. Deterministic Models, Stochastic Models, Statistical models and State Space Models are the main four categories of statistical modelling. Number of susceptible individuals, infected individuals and number of AIDS cases used as parameters in Deterministic modeling. But in Stochastic models, it will be random variables and superior than deterministic models. Statistical models are better than stochastic and deterministic because it uses the epidemiological and survey data with back calculation methodology. HIV incubation period is the random time between the HIV-infection and the onset of clinical AIDS. Distribution of this non-negative random variable is known as HIV incubation period distribution. Back-calculation method reconstructs the past pattern of HIV infection and predicts the future number of AIDS cases with the present infection status. It depends on three important factors: incubation distribution, incidence curve and observed number of AIDS cases over time. This method is very popular and requires less information and assumptions. Lack of information about incubation distribution, the effect of intervention therapy on incubation period, and errors in reported AIDS incidence leads to uncertainties associated with this method. The incubation distribution is assumed to be exactly known in back-calculation methodology. Incubation period of HIV is very long and highly variable within and between cohorts. The current prevalence of HIV-infection and the corresponding pattern of incidence from the beginning of the epidemic to the present time are mainly estimated by means of back-calculation method. It calculates the most likely temporal distribution of infected individuals compatible with the number of observed AIDS cases starting from the suitable estimate of the incubation period, derived from the available data. State space models have the combined effect of stochastic and statistical models but it is mainly for engineering data and used in forecasting of AIDS. There are also uncertainties associated with these approaches. In addition to this the recent advance in this field used is curve fitting models. Sathian and Sreedharan used this method for forecasting several infectious and non infectious diseases. It gives more accurate estimates compared to the other models [2-28].

Curve fitting method

The Curve Estimation procedure produces curve estimation regression statistics and related plots for 11 different curve estimation regression models. A separate model is produced for each dependent variable. You can also find out predicted values, residuals, and prediction intervals as new variables. For each model, we can find out the regression coefficients, multiple R, R², adjusted R², standard error of the estimate, analysis-of-variance table, predicted values, residuals, and prediction intervals as new variables. Models: linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth, and exponential.

Linear Model

Model whose equation is \( Y = b_0 + (b_1 \times t) \). The series values are modeled as a linear function of time.

Logarithmic Model

Model whose equation is \( Y = b_0 + (b_1 \times \ln(t)) \).

Inverse Model

Model whose equation is \( Y = b_0 + (b_1 / t) \).
Quadratic Model
Model whose equation is \( Y = b_0 + (b_1 \cdot t) + (b_2 \cdot t^2) \). The quadratic model can be used to model a series that "takes off" or a series that damps.

Cubic Model
Model that is defined by the equation \( Y = b_0 + (b_1 \cdot t) + (b_2 \cdot t^2) + (b_3 \cdot t^3) \).

Power Model
Model whose equation is \( Y = b_0 \cdot (t^{b_1}) \) or \( \ln(Y) = \ln(b_0) + (b_1 \cdot \ln(t)) \).

Compound Model
Model whose equation is \( Y = b_0 \cdot (b_1^t) \) or \( \ln(Y) = \ln(b_0) + (\ln(b_1) \cdot t) \).

S-curve Model
Model whose equation is \( Y = e^{(b_0 + (b_1/t))} \) or \( \ln(Y) = b_0 + (b_1/t) \).

Logistic Model
Model whose equation is \( Y = 1 / (1/u + (b_0 * (b_1^{t})) \) or \( \ln(1/y-1/u) = \ln(b_0) + (\ln(b_1) \cdot t) \) where \( u \) is the upper boundary value. After selecting Logistic, specify the upper boundary value to use in the regression equation. The value must be a positive number that is greater than the largest dependent variable value.

Growth Model
Model whose equation is \( Y = e^{(b_0 + (b_1 \cdot t))} \) or \( \ln(Y) = b_0 + (b_1 \cdot t) \).

Exponential Model
Model whose equation is \( Y = b_0 \cdot (e^{b_1 \cdot t}) \) or \( \ln(Y) = \ln(b_0) + (b_1 \cdot t) \).

The annual numbers of HIV patients will be plotted in y-axis against the corresponding year in the x-axis. Curve fitting, also known as regression analysis, will be used to find the "best fit" line or curve for a series of data points. F-test should be used for selecting the best fitting curve for the testing of hypothesis. P-value must be taken as significant when < 0.05 (two-tailed). \( R^2 \) value > 0.80 should be taken as significantly better for prediction. The decision regarding the selection of a suitable prediction approach is governed by the relative performance of the models for monitoring and prediction. It should also adequately interpret the phenomenon under study.

Conclusion
This paper has reviewed a novel method of using the curve fitting method in HIV/AIDS data. The approach is simple to understand and apply, and is capable of curve fitting a whole range of different models. It also has the advantage that several different models, for a given data series, can be easily investigated, thus easing the model selection dilemma.

References

14. Brookmeyer R and Gail MH. AIDS epidemiology:


Disclaimer

This article has been downloaded from WebmedCentral. With our unique author driven post publication peer review, contents posted on this web portal do not undergo any prepublication peer or editorial review. It is completely the responsibility of the authors to ensure not only scientific and ethical standards of the manuscript but also its grammatical accuracy. Authors must ensure that they obtain all the necessary permissions before submitting any information that requires obtaining a consent or approval from a third party. Authors should also ensure not to submit any information which they do not have the copyright of or of which they have transferred the copyrights to a third party.

Contents on WebmedCentral are purely for biomedical researchers and scientists. They are not meant to cater to the needs of an individual patient. The web portal or any content(s) therein is neither designed to support, nor replace, the relationship that exists between a patient/site visitor and his/her physician. Your use of the WebmedCentral site and its contents is entirely at your own risk. We do not take any responsibility for any harm that you may suffer or inflict on a third person by following the contents of this website.