Statistical Modelling and Forecasting of Reported HIV Cases in Nepal

Sathian B[1], Sreedharan J[2], Mittal A[3], Baboo NS[4], Chandrasekharan N[5], Devkota S[1], Abhilash ES[6], Rajesh E[7], Dixit SB[8]

[1]Assistant Professor, Department of Community Medicine, Manipal College of Medical Sciences, Pokhara, Nepal

[2]Assistant Director (Research), Gulf Medical University, Ajman, United Arab Emirates

[3]Associate Professor, Department of Biochemistry, Manipal College of Medical Sciences, Pokhara, Nepal.

[4] Professor, Department of Physiology, Manipal College of Medical Sciences, Pokhara, Nepal

[5]Assistant Professor, Department of Orthopaedics, Manipal College of Medical Sciences, Pokhara, Nepal.

[6]Lecturer, Department of Biological Sciences, Grammar School (UAE).

[7]Assistant Professor, School of Health and Behavioural Sciences, Mahatma Gandhi University, India.

[8] Professor & Head, Department of Community Medicine, Manipal College of Medical Sciences, Pokhara, Nepal.

Original Article

Corresponding Author:
Dr. Brijesh Sathian,
Assistant Professor, Department of Community Medicine,
Manipal College of Medical Sciences
Deep Heights 16, PO BOX 155, Pokhara, Nepal.
Email:*drsathian@gmail.com*

Abstract

Background

The real state about the spread of the HIV epidemic in Nepal is not clear since the details available are on the basis of risk group. The objective of the study is to extract as much as information possible from available data and find out the trends of HIV cases in future.

Material and methods

A retrospective study was carried out on the data collected from the Health ministry records of Nepal, between 1988 and 2004. Descriptive statistics and statistical modelling were used for the analysis and forecasting of data.

Results

Excluding the constant term from the equation, the cubic model was the best fit, for the forecasting of HIV cases. NCASC reported cumulative number of HIV cases up to 2009 differs from our projected cases by 46 (99.99% accuracy in prediction). Using cubic equation, it is estimated that 4773 males, 2163 females and 6936 total reported number of HIV cases will be there in Nepal by the year 2015.

Conclusion

The HIV cases in Nepal are having an increasing trend. Estimates of the total number of prevalent HIV infections attributable to the major routes of infection make an important contribution to public health policy. They can be used for the planning of healthcare services and for contributing to estimates of the future numbers with severe HIV infection used for planning health promotion programmes.

Key words: Statistical Modelling, HIV, Nepal

Background

The current situation of HIV in Nepal is different from when

Nepal Journal of Epidemiology 2011;1(3): 106-110

Published online by NepJOL-INASP
www.nepjol.info/index.php/NJE

the first case was diagnosed in 1988. Up until recently, (2009) the total number of positive cases reported are 14787 out of which 13005 are receiving HIV care. There are gaps and challenges to be addressed in the fight against HIV and AIDS. Nepal is low prevalence country for HIV and AIDS (0.49 percent). However, some of the groups show evidence of a concentrated HIV epidemic e.g. sex workers, migrant population and intravenous drug users (IVDU's), both in rural and urban areas. Since 1988 when the first case was diagnosed Ministry of Health and Population/Department of Health Service (MoHP/DoHS) and different stakeholders came forward to address HIV and AIDS issues. The main focus was given to preventive aspects. In 1995 MoHP in consultation with different stakeholders developed a policy for the control of HIV and AIDS. However, the activities were implemented in a sporadic and disorganized manner.

The real state about the spread of the epidemic in Nepal is not clear since the details available are on the basis of risk group. As regards the risk group, the prevalence rate is high. Perhaps it may not represent it prevalence rate of the general population[1,2]. The study conducted by Kermack and McKendrick (1927) for treatment of the Bombay plague of 1905–06 proved the capability of mathematical models in understanding and predicting epidemics. Anderson and May (1991) present more models of infections including HIV with illustrations. Mukerji (1989) represents one of the earliest Indian attempts at modeling data on AIDS[3-5]. This model used annualized south Asian regional data and extrapolated to AIDS in future. Basu et al (1998) attempt to model the spread of AIDS in a comprehensive manner with limited data[6]. The applicability of various models to predict AIDS in India, beginning from classic simple epidemic models to more complex heterosexual transmission models proposed and back calculation method were done by Sreenivasa Rao (2003)[7]. Williams (2005), et al studied about HIV prevalence in India. Joshua (1999), et al studied methods for modeling the HIV/AIDS epidemic in Sub Saharan Africa[8,9]. Until the last decade conventional study through statistical methods were adopted to understand the trend and prevalence of HIV/AIDS in almost all countries. The objective of the study is to extract as much as information possible from available data and find out the trends of HIV cases in future.

Materials and Methods

A retrospective study was carried out on the HIV data collected from the Health ministry records of Nepal, between 1988 and 2004. The major mode of transmission of HIV in the country is heterosexual. The numbers compiled on the basis of reported voluntary testing and sentinel surveillance. The data was analysed using Excel 2003, R 2.8.0, Statistical Package for the Social Sciences (SPSS) for Windows Version 16.0 (SPSS Inc; Chicago, IL, USA) and EPI Info 3.5.1 windows version. A p-value of < 0.05 (two-tailed) was used to establish statistical significance. The annual reported numbers of HIV patients plotted in y-axis against the corresponding year in the x-axis. Curve fitting, also known as regression analysis, was used to find the "best fit" line or curve for a series of data points. Linear, Logarithmic, Inverse, Quadratic, and Cubic were chosen to fit to the obtained curve. F-test was used for selecting the best fitting curve for the testing of hypothesis. P-value was taken as significant when < 0.05 (two-tailed). R2 value > 0.80 was taken as significantly better for prediction[10]. 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. Cubic model selected here could closely fit curves for estimated and reported HIV cases (Fig 1). While building model, the extremities (maximums and minimums) play a great role. If the points are scattered more, the curve tries to adjust with maximum number of observed points. The cubic model is a third degree polynomial, represented by the equation $y = m_0 + m_1 * x + m_2 * x^2 + m_3 * x^3$, where m_0 is the constant term and m_1, m_2, m_3 are coefficient terms[11,12]. Without the constant term, the equation of this model is $y = m_1 * x + m_2 * x^2 + m_3 * x^3$. Where Y is the number of number of reported HIV cases annually and X is the corresponding year; 1=1988, 2=1989, 3=1990, 4=1991 and so on.

Results

The data was modelled using the curve fitting method. [Tables 1, 3 and 5/Graph 1] depicts the model summary and the parameter estimates including and excluding the constant term for different models. When the constant term was included, the p values were >0.05 in all the models and none of the models were best fitted. After excluding the constant term from the equation, the cubic model was the best fit, for the forecasting of HIV cases. The cubic model equations below (1, 2 and 3) contain X and Y, which are the corresponding year and frequency of reported HIV cases respectively. m_1, m_2, m_3 calculated from the observed data.

Table 1: Model summary and parameter estimates excluding the constant term for different models for the reported number of male HIV cases

Equation	Model Summary	
	R Square	p-value
Linear	0.734	0.001
Logarithmic	0.600	0.001
Inverse	0.026	0.525
Quadratic	0.882	0.001
Cubic	0.908	0.001

The equation for the cubic model for the reported number of male HIV cases is

$$Y= 25.11X - 5.029X^2 + 0.365X^3 \text{ --------- (1)}$$

(Where Y is the number of number of reported male HIV cases annually and X is the corresponding year; 1=1988, 2=1989, 3=1990, 4=1991 and so on)

Published online by NepJOL-INASP
www.nepjol.info/index.php/NJE

Using the equation (1), reported numbers of male HIV cases were estimated.

The equation for the cubic model for the reported number of female HIV cases is

$$Y = 20.768X - 3.756X^2 + 0.206X^3 \text{ --------- (2)}$$

(Where Y is the number of number of reported female HIV cases annually and X is the corresponding year; 1=1988, 2=1989, 3=1990, 4=1991 and so on)

Table 2: Reported number of HIV cases up to the year 2004, and estimated number of HIV cases up to the year 2015.

Year	Estimated cases	Predicted cases	Lower Limit	Upper Limit
1988	3	20	0	249
1989	0	33	0	269
1990	2	40	0	281
1991	12	43	0	287
1992	39	45	0	288
1993	41	48	0	289
1994	18	55	0	293
1995	71	66	0	303
1996	50	85	0	323
1997	394	113	0	353
1998	166	154	0	395
1999	174	208	0	451
2000	301	279	36	521
2001	264	368	126	609
2002	360	477	235	720
2003	505	610	357	863
2004	942	767	484	1050
2005		951	611	1292
2006		1165	733	1598
2007		1411	852	1970
2008		1690	969	2411
2009		2005	1087	2923
2010		2358	1206	3511
2011		2752	1327	4177
2012		3188	1450	4926
2013		3669	1575	5763
2014		4197	1703	6691
2015		4773	1832	7714

Table 3: Model Summary and Parameter Estimates excluding the constant term for different models for the reported number of female HIV cases

Equation	Model Summary	
	R Square	p-value
Linear	0.740	0.001
Logarithmic	0.628	0.001
Inverse	0.035	0.456
Quadratic	0.846	0.001
Cubic	0.909	0.001

Using the equation (2), reported numbers of female HIV cases were estimated. Table 4 shows the reported number of HIV cases up to the year 2004, and estimated number of HIV cases up to the year 2015.

The equation for the cubic model for the reported number of total HIV cases is

$$Y = 45.879X - 8.784X^2 + 0.571X^3 \text{ --------- (3)}$$

(Where Y is the number of number of reported total HIV cases annually and X is the corresponding year; 1=1988, 2=1989, 3=1990, 4=1991 and so on)

Table 4: Observed and estimated HIV cases of females

Year	Observed cases	Estimated cases	Lower Limit	Upper Limit
1988	1	17	0	98
1989	2	28	0	112
1990	3	34	0	119
1991	14	36	0	122
1992	38	36	0	122
1993	40	34	0	119
1994	22	32	0	116
1995	39	31	0	115
1996	85	33	0	117
1997	95	38	0	123
1998	54	48	0	134
1999	48	65	0	151
2000	95	88	2	174
2001	60	120	35	206
2002	107	162	77	248
2003	209	215	126	305
2004	340	281	181	381
2005		359	239	480
2006		453	300	606
2007		563	365	760
2008		689	434	944
2009		835	510	1159
2010		1000	592	1407
2011		1185	681	1690
2012		1394	779	2008
2013		1625	884	2366
2014		1881	999	2763
2015		2163	1123	3203

Table 5: Model Summary and Parameter Estimates excluding the constant term for different models for the reported number of total HIV cases

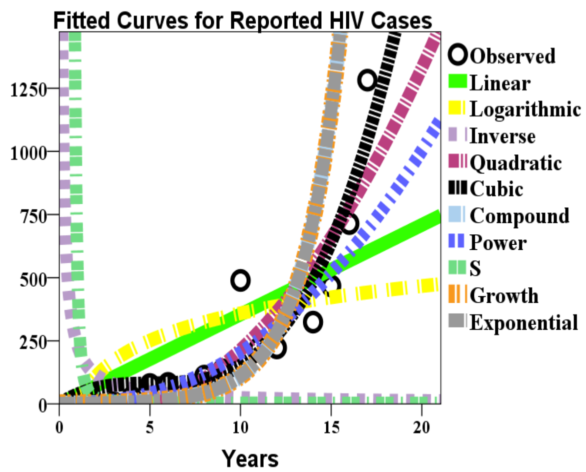
Equation	Model Summary	
	R Square	p-value
Linear	0.743	0.001
Logarithmic	0.613	0.001
Inverse	0.028	0.504
Quadratic	0.881	0.001
Cubic	0.915	0.001

Using the equation (3), reported numbers of total HIV cases were estimated. Table 6 shows the reported number of HIV cases up to the year 2004, and estimated number of HIV cases up to the year 2015.

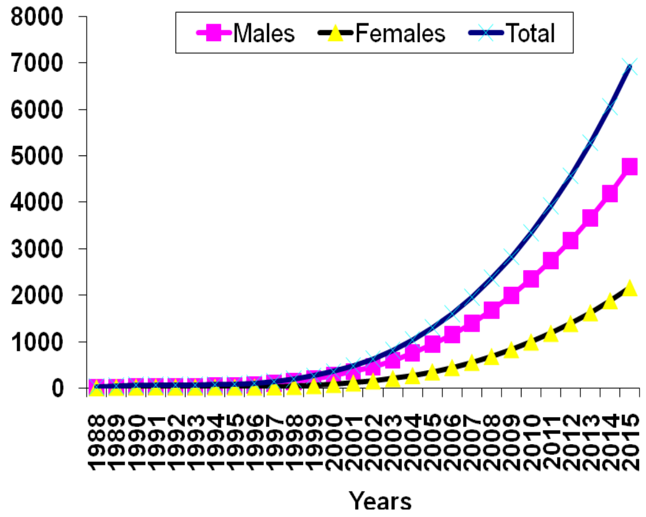
Table 6: Observed and estimated Total HIV cases

Year	Observed cases	Estimated cases	Lower Limit	Upper Limit
1988	4	38	0	334
1989	2	61	0	367
1990	5	74	0	386
1991	26	80	0	394
1992	377	81	0	395
1993	81	82	0	393
1994	40	87	0	395
1995	110	97	0	404
1996	135	118	0	426
1997	489	152	0	462
1998	220	202	0	515
1999	222	273	0	587
2000	396	367	53	681
2001	324	488	176	800
2002	467	639	326	953
2003	714	825	498	1152
2004	1282	1048	682	1413
2005		1311	870	1752
2006		1618	1059	2178
2007		1973	1250	2697
2008		2379	1447	3312
2009		2840	1652	4028
2010		3358	1867	4849
2011		3937	2093	5782
2012		4582	2332	6831
2013		5294	2584	8003
2014		6078	2851	9305
2015		6936	3131	10742

Graph 1: Fitted curves for reported HIV cases
(X-axis shows years; 1=1988, 2=1989, 3=1990, 4=1991 and so on, Y-axis shows number of reported HIV cases)



Graph 2: Year wise estimates for reported HIV cases.



Discussion

<u>Modelling and Extrapolation:</u> A plot is a graphical representation of the collected data (independent and dependent variables) involved in a study. The association between these variables are then assessed by connecting the `points' with a line. Though very true, this association cannot be relied upon to predict the future trend of this data. Now a `model', which `fits best' to the observed data has to be worked out. This is then `fitted' and used to replace the existing set of data points as `the appropriate model'. After `modelling' the observed data, this model can be used to predict future trend of the dependent variable for a given change in the other. The foregoing statement covertly mentions several requirements which often ensure confident achievement in any subsequent extrapolation from the model. The model selected must be the most appropriate for the collected data. A usable and understandable curve-fitting method is to be available from which the model facts those are reflective of future behaviour can be obtained [13,14].

Timely and accurate monitoring of the HIV epidemic requires measures of incidence, that is, the number of new infections in a defined population that occur during a defined time period. Unfortunately, longitudinal studies that have traditionally provided incidence measures are costly, time consuming, logistically complex, and may be subjectively biased, differential loss to follow-up, or an intervention effect [15-17]. As a result, public health agencies have generally relied on surveys that measure HIV prevalence, the proportion of persons at a specified point in time that are infected, to track the epidemic[18].

Using the curve fitting method, we estimated the number and trend of reported HIV cases at Nepal from the year 1988 to 2015. Cubic model provided closely fitted curves for estimated and reported HIV cases (Graph 1). While building model, the extremities (maximums and minimums) play a great role. If the points are scattered more, the curve tries to adjust with maximum number of observed points. Therefore, it might give over- and under-estimation inevitably, but that is not the case in all the situations. A sudden annual decrease and increase in the trend is possible, as the curve cannot exactly connect these data points because of its shape. For adjusting the over-and under-estimation, the model gave wide confidence intervals in case of some years (Table 6). In our study, the future annual reported HIV cases (Table 6) shows an increasing trend. Such an increase might be convincing as HIV incidence in developing countries is expected to rise principally due to the possible decline of mortality from infectious diseases, population growth. As of 2007, national estimates indicate that approximately 70,000 adults and children are infected with the HIV virus in Nepal, with an estimated prevalence of about 0.49% in the adult population. As of June 2007, a total of 9756 cases of HIV, 1454 AIDS cases and 423 AIDS deaths had been reported to the National Centre for AIDS and STD control (NCASC). According to our study, we got cumulative number of Reported HIV cases 9614 up to 2007 that is just 142 cases lesser than NCASC reported. Like wise up to 2009 projected is 14833 which differs with actual reported cumulative number of HIV cases by 46 cases[1,2]. Our study hereby establishes the applicability of statistical modelling in predicting the reported number of HIV cases in the Nepalese context.

Conclusion

It is well known that HIV/AIDS is a fatal epidemic and effective medicines are not yet discovered hence active precautionary measures have to be taken with intensive care against the spread of the epidemic. In order to determine the current levels and trends in this epidemic, best possible information is extremely necessary. The correct information on incidence and prevalence is not possible now due to stigma of the disease. The use of statistical modeling approaches make a valuable contribution in order to develop better understanding of the levels and trends in the HIV epidemic and the limited information based on the estimates. Estimates of the total number of prevalent HIV infections attributable to the major routes of infection make an important contribution to public health policy. They can be used for the planning of healthcare services and for contributing to estimates of the future numbers with severe HIV infection used for further planning of the programmes[19].

References

1. Annual Report, Department of Health Service, Ministry of Health and Population.2007;179-83.
2. Fact Sheet: HIV and AIDS Epidemic Update of Nepal. National Center for AIDS and STD Control, Ministry of Health and Population. November 2009.
3. Kermack WO and McKendrick AG Proc. R. Soc. London. 1927;A115:700–21.
4. Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control. Oxford, UK: Oxford University Press; 1991. 768pp.
5. Mukerji S. Dynamics of population and family welfare. (Eds. Srinivasan K and Pathak KB). New Delhi: Himalaya Publishing House, 1989. pp. 300-14.
6. Basu A, Basu S, Chakraborty MS, Dewanji A, Ghosh JK, Majumder PP. Projection of HIV infection in Calcutta. Indian J Med Res. 1998;107:159-72.
7. Rao S, Mathematical modelling of AIDS epidemic in India. Current Science. 2003; 84(9):1192-97.
8) Williams BG, Granich R, Chauhan LS, Dharmshaktu NS, Dye C. The impact of HIV / AIDS on the control of tuberculosis in India. Proceedings of the National Academy of Sciences of the United States of America. 2005 Jul 5;102(27):9619-24.
9) Salomon JA, Gakidou EE, Murray CJL. Methods for modelling the HIV/AIDS Epidemic in Sub-Saharan Africa. GPE Discussion Paper Series: No:3.1999. http://www.who.int/entity/healthinfo/ paper03.pdf.
10) Petrie A, Sabin C. Medical Statistics at a glance. I Ed, London, Blackwell Science. 2000. pp 71-74.
11) Beyer WH. Standard Mathematical Tables. Cleveland, CRC Press. 1976.
12) Chambers JM, Cleveland WS, Kleiner B., et al. Graphical Methods for Data Analysis. Boston, Duxbury Press. 1983.
13) Hindmarch GW and Towill DR. Theory and Application of the Time Constant Learning Curve Model. ORSA/TIMS, Puerto Rico; October 1974.
14) Nie NH. SPSS: Statistical Package for the Social Sciences. New York: McGraw-Hill; 1975.
15. Brookmeyer R, Quinn T, Shepherd M, Mehendale S, Rodrigues J, Bollinger R. The AIDS epidemic in India: a new method for estimating current human immunodeficiency virus (HIV) incidence rates. Am J Epidemiol 1995; 142:709-13.
16) Heyward W, Osmanov S, Saba J, et al. Preparation for phase III HIV vaccine efficacy trials: methods for the determination of HIV incidence. AIDS 1994; 8:1285-92.
17) Laga M, Alary M, Nzila N, et al. Condom promotion, sexually transmitted diseases treatment, and declining incidence of HIV-1 infection in female Zairian sex workers. Lancet 1994; 344:246-8.
18) U.S. Centers for Disease Control and Prevention. National HIV serosurveillance summary: results through 1992, vol 3. Atlanta: US Department of Health and Human Services 1994:1-51.
19. Gill ON, Adler MW, Day NE. Monitoring the prevalence of HIV: foundations for a programme of unlinked anonymous testing in England and Wales. BMJ 1989;299:1295–8.