ORIGINAL ARTICLE

Forecasting the Number of Human Immunodeficiency Virus Infections in the Korean Population Using the Autoregressive Integrated Moving Average Model

Hye-Kyung Yu [a], Na-Young Kim [a], Sung Soon Kim [a], Chaeshin Chu [b], Mee-Kyung Kee [a,*]

[a]*Division of AIDS, Korea National Institute of Health, Osong, Korea.*
[b]*Division of Epidemic Intelligence Service, Korea Centers for Disease Control and Prevention, Osong, Korea.*

Abstract

Objectives: From the introduction of HIV into the Republic of Korea in 1985 through 2012, 9,410 HIV-infected Koreans have been identified. Since 2000, there has been a sharp increase in newly diagnosed HIV-infected Koreans. It is necessary to estimate the changes in HIV infection to plan budgets and to modify HIV/AIDS prevention policy. We constructed autoregressive integrated moving average (ARIMA) models to forecast the number of HIV infections from 2013 to 2017.

Methods: HIV infection data from 1985 to 2012 were used to fit ARIMA models. Akaike Information Criterion and Schwartz Bayesian Criterion statistics were used to evaluate the constructed models. Estimation was via the maximum likelihood method. To assess the validity of the proposed models, the mean absolute percentage error (MAPE) between the number of observed and fitted HIV infections from 1985 to 2012 was calculated. Finally, the fitted ARIMA models were used to forecast the number of HIV infections from 2013 to 2017.

Results: The fitted number of HIV infections was calculated by optimum ARIMA (2,2,1) model from 1985—2012. The fitted number was similar to the observed number of HIV infections, with a MAPE of 13.7%. The forecasted number of new HIV infections in 2013 was 962 (95% confidence interval (CI): 889—1,036) and in 2017 was 1,111 (95% CI: 805—1,418). The forecasted cumulative number of HIV infections in 2013 was 10,372 (95% CI: 10,308—10,437) and in 2017 was 14,724 (95% CI: 13,893—15,555) by ARIMA (1,2,3).

Conclusion: Based on the forecast of the number of newly diagnosed HIV infections and the current cumulative number of HIV infections, the cumulative number of HIV-infected Koreans in 2017 would reach about 15,000.

*Corresponding author.
E-mail: keemeekyung@gmail.com

1. Introduction

Human immunodeficiency virus (HIV) infection can show nonspecific symptoms in the early stage such as anorexia, weight loss, fever, lymph node enlargement, and fatigue, but has a long asymptomatic period after an acute infection period. Therefore, the real prevalence of HIV infection has been estimated to be more than the reported numbers. Since the first reported case of AIDS, the United Nations Programme on HIV/AIDS (UNAIDS) estimated the global people living with HIV/AIDS (PLWHA) to be 35.3 million in 2012 [1]. The cumulative number of HIV-infected people identified in the Republic of Korea from 1985 to 2012 was 10,453 (9,410 Koreans, 1,043 foreigners), relatively low compared to many other countries [2]. However, since 2000, the number of the HIV infected people in Korea has sharply increased; 800 to 1,000 persons are newly identified each year in the recent 3 years [2], and it is questionable whether this increasing trend will continue.

The United States stated the research on estimation in 1980, and many countries have estimated the HIV-infected persons using the model suggested by UNAIDS for identifying the trend of their own nation. The method to estimate incidence, prevalence, the number of infected persons, and the number of deaths has been improved since its development in late 1980. It includes the back calculation method, ratio method, Delphi survey method, mathematical and computer/simulation models, and Epimodel. No single optimal estimation method exists; each method has its own characteristics. Recent standard methods for estimation and prediction are the workbook method, estimation and projection package (EPP) method, and time series method (autoregressive moving average model).

As for the estimation of HIV-infected Koreans, the back calculation method, Epi-info model, and EPP model cannot be applied because of limitations including: the lack of exact statistics on AIDS cases and no specific information available for commercial sex workers (CWSs), intravenous drug users (IDUs), and the sexually transmitted infection risk group (STI RG). However, the autoregressive integrated moving average (ARIMA) time series method utilizes autocorrelation and can produce an estimation model with information on HIV cases identified yearly or monthly. ARIMA is used for the estimation of influenza mortality [3], malaria incidence [4], and other infectious diseases [5–7].

This study aimed to construct estimation models using the annual number of newly identified HIV-infected Koreans in Korea from 1985 to 2012, and to forecast the number of HIV cases and the trend in epidemiological characteristics.

2. Materials and Methods

2.1. Materials

This study has utilized 9410 identified HIV-infected persons' basic epidemiological data, including diagnosed year, sex, age, screening site, and date of death from 1985 to 2012 in Korea. The local public health centers conduct an epidemiological investigation on the route of infection once the HIV-infected persons are identified and report the results to the Korea Centers for Disease Control and Prevention (KCDC). The epidemiological data are managed by the KCDC HIV database system. The data for this study were collected from the KCDC HIV database system.

2.2. Methods

ARIMA models are the most commonly used time series prediction models. We constructed ARIMA models with data from HIV cases in Korea from 1985–2012, to forecast the annual number of newly diagnosed HIV-infected Koreans and to forecast the cumulative number of HIV cases in the next 5 years (2013–2017). To determine the degree of ARIMA, we have utilized the autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs and constructed an optimal model on the basis of Akaike's information criterion (AIC) statistic and Schwartz's Bayesian criterion (SBC) statistic. Conditional least squares method (CLS) was used for parameter estimation.

This study constructed models by sex, age, and screening site, similar to that of the study by Debanne et al [8]; as a result, the sums of the forecasting numbers by variables are not equal to the forecasting number of newly diagnosed HIV cases. Forecasting numbers were calculated based on a percentage of the forecasting value of each category by each variable. To check the accuracy of each model, mean absolute percentage errors (MAPE) between the number of observed and fitted HIV infections from 1985 to 2012 were calculated. Finally, the number of HIV-infected Koreans in each year from 2013–2017 were forecasted. SAS version 9.3 (SAS Institute, Cary, NC, USA) was used for data analysis.

3. Results

3.1. Status of HIV-infected persons in Korea until 2012

Figure 1 depicts the number of people with HIV in Korea from 1985 to 2012. The cumulative number of HIV-infected Koreans through 2012 is 9,410 (8,668 men, 742 women), and the number of PLWHA is 7,788 (7,165 men, 624 women). The number of HIV case sharply increases from 2000 to 2008. However, the past 5 years have shown a lower increase of newly diagnosed cases.
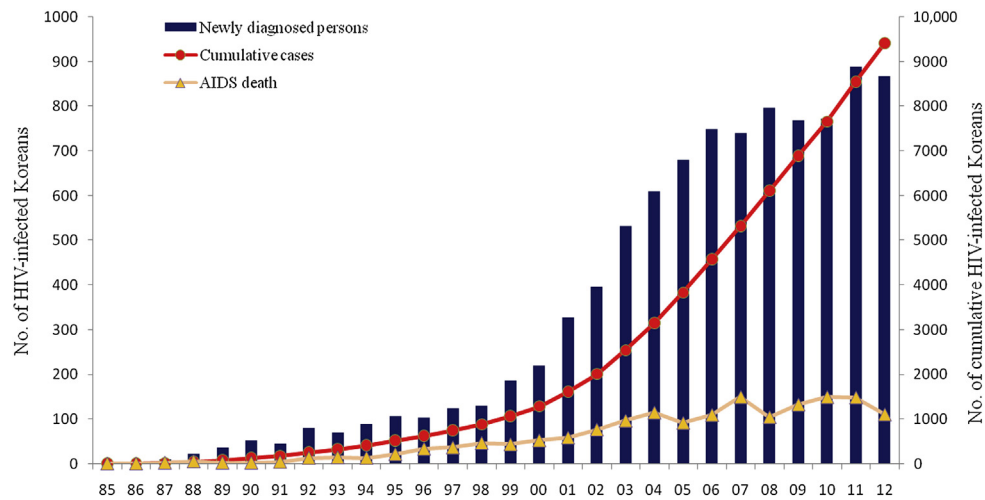
Figure 1. The trend in the number of HIV-infected Koreans, 1985−2012. The bars are the number of newly diagnosed HIV-infected Koreans, the dotted red line is the cumulative number of HIV-infected Koreans, and the yellow triangle line is the number of AIDS deaths per year.

3.2. Construction of models to estimate the number of HIV-infected persons, and forecast of 2013−2017

Figure 2 represents observed yearly data (1985−2012), predicted values, and 95% confidence intervals (CI) for the predicted values. Most of observed data are contained within the 95% CI of the predicted value, indicating that the fitted models are reasonable.

The forecasts of each ARIMA model for the number of HIV-infected Koreans in 2013−2017 are shown in Table 1. Each model shows the number of HIV-infected Koreans increases each year, so the data were stabilized by
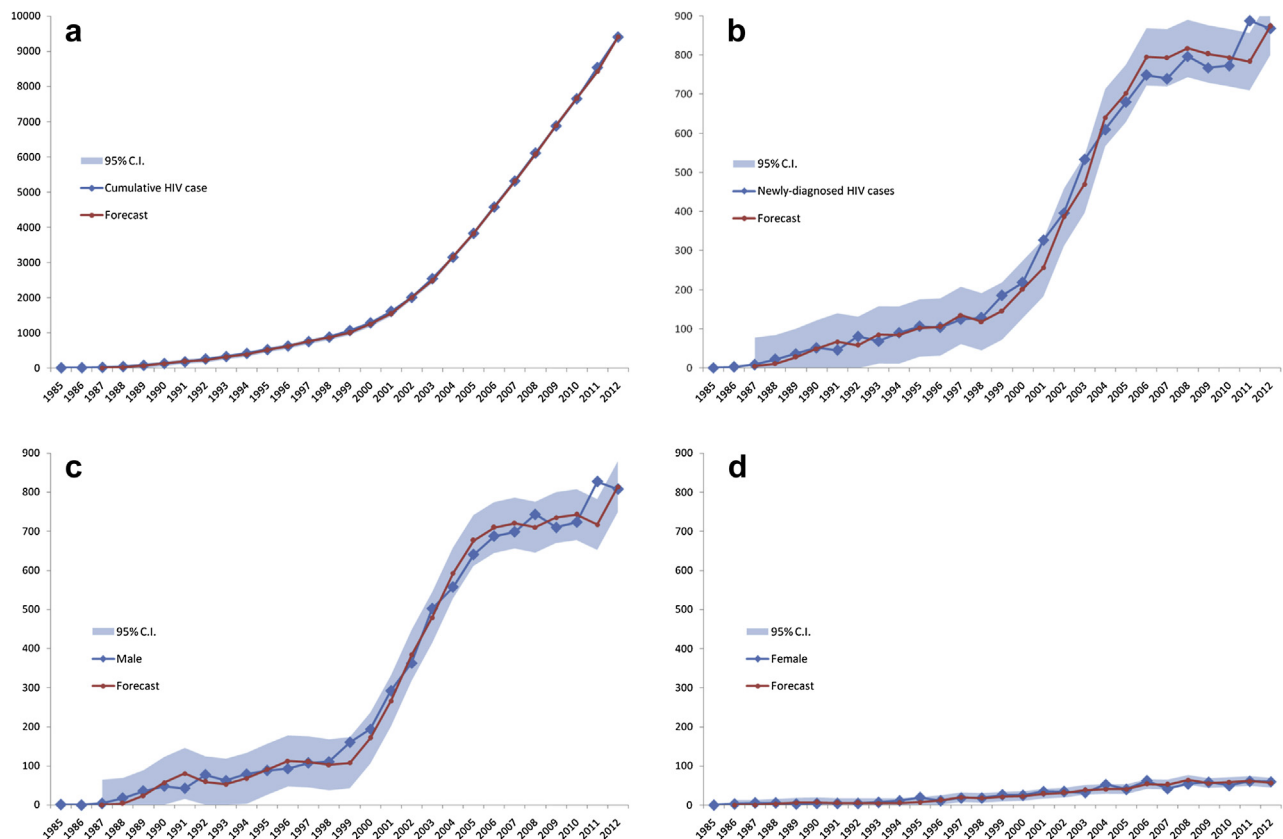

Figure 2. Comparison of the reported and forecasted numbers of HIV cases in Korea from 1985 through 2012. Diamond line: actual number, dot line: forecasted number, band: 95% confidence interval of forecasted number. (a) The cumulative number of HIV-infected Koreans, (b) the number of newly diagnosed HIV-infected Koreans, (c) the number of newly diagnosed HIV infected male Koreans, and (d) the number of newly diagnosed HIV infected female Koreans.

Table 1. The forecasted number of HIV-infected Koreans in Korea in 2013−2017

Characteristics	Model	Forecast				
		2013	2014	2015	2016	2017
Cumulative HIV cases	ARIMA (1,2,3)	10,372 (10,308−10,437)	11,425 (11,275−11,574)	12,470 (12,170−12,770)	13,595 (13,058−14,131)	14,724 (13,893−15,555)
People living with HIV	ARIMA (1,2,2)	8,615 (8,533−8,698)	9,508 (9,325−9,691)	10,438 (10,075−10,802)	11,391 (10,768−12,013)	12,355 (11,404−13,306)
Newly-diagnosed HIV cases	ARIMA (2,2,1)	962 (889− 1,036)	992 (882−1,102)	1,026 (856−1,196)	1,108 (867−1,349)	1,111 (805−1,418)
Sex						
Man	ARIMA (1,2,2)	902 (838−966)	934 (845−1,022)	969 (811−1,127)	1050 (801−1,299)	1055 (723−1,388)
Woman	ARIMA (2,1,2)	60 (48−72)	58 (47−70)	57 (43−70)	58 (43−73)	56 (39−72)
Age						
<20	ARIMA (1,2,1)	44 (37−51)	42 (34−51)	51 (39−64)	54 (39−69)	56 (38−74)
20−29	ARIMA (1,1,1)	302 (260−344)	318 (254−382)	369 (276−463)	401 (282−521)	414 (273−554)
30−39	ARIMA (1,2,0)	205 (169−242)	200 (143−258)	212 (111−312)	214 (73−355)	204 (23−385)
40−49	ARIMA (2,2,0)	185 (151−218)	203 (161−245)	155 (96−213)	188 (96−279)	190 (85−296)
50−59	ARIMA (2,2,0)	148 (126−171)	153 (123−183)	163 (117−208)	173 (107−239)	172 (92−251)
≥60	ARIMA (2,2,0)	78 (58−97)	76 (44−108)	76 (31−122)	78 (13−143)	75 (0−157)
Screening site						
Public institution	ARIMA (2,1,2)	145 (99−192)	134 (75−193)	135 (65−205)	153 (65−241)	164 (60−269)
Hospital	ARIMA (1,1,2)	714 (628−799)	743 (626−860)	780 (600−960)	829 (568−1089)	830 (499−1161)
Blood bank	ARIMA (1,1,1)	103 (82−125)	115 (85−144)	111 (72−149)	126 (81−173)	117 (65−169)

difference and fitted. The MAPE for the cumulative number of HIV-infected Koreans is 5.5%, for the cumulative number of PLWHA is 6.3%, and for the number of newly-diagnosed HIV-infected Koreans is 13.7%. By 2017, these models estimate that Korea will have a cumulative number of HIV-infected Koreans of 14,724 (95% CI; 13,893−15,555), and a cumulative PLWHA of 12,355 (95% CI: 11,404−13,306). A total of 1,111 (95% CI: 805−1,418) Koreans are estimated to be newly diagnosed with HIV in 2017. Analysis by epidemiological characteristic shows that the difference between men and women will be greater in 2017: 1,288 men (95% CI: 882−1,693) vs. 68 women (95% CI: 48−87). The age group in which the largest number of HIV-infected Koreans is expected to be diagnosed is 20−29, with an average of 387 persons in their twenties (95% CI; 255−518) being newly diagnosed per year. HIV-infected Koreans are expected to be mostly diagnosed in hospitals, followed by public health centers and then blood banks.

4. Discussion

Models reveals that by 2017 the cumulative number of HIV-infected Koreans will be 14,724, with 12,355 cumulative PLWHA, and 1,111 newly diagnosed persons in the year 2017. It took 29 years since the first diagnosis of HIV infection in 1985 to identify a cumulative number of 10,000 HIV-infected people in Korea. This study suggests that the cumulative trend will increase dramatically, to 14,724 in 2017 and over 15,000 in 2018, for an increase of 5,000 in 6 years. The number of newly diagnosed HIV-infected Koreans is also expected to increase, with more than 1,000 per year after 2014.

The current gender difference (11:1 men to women) is expected to widen even further; the gender ratio of newly diagnosed HIV-infected Koreans will be 19:1 in 2017. By 2012, the age group with the highest cumulative number of HIV-infected Koreans was 30−39 (29%), but there is also a recent increase of HIV-infected Koreans in their twenties. Based on the results of this study, we can expect a sharp increase of HIV-infected Koreans in their twenties for the upcoming 5 years. Increases in the number of teens and those in their fifties are also forecasted. This result suggests that a main target group for prevention should be those aged 20−29 in Korea. In previous studies on STIs prevalence (HIV and HSV-2), the prevalence were low in Koreans in their twenties [9−11], but it is necessary to monitor this age group carefully for any changes.

This study has at least one limitation: the fitted models using the number of HIV-infected Koreans during 1985−2012 assume that past HIV infection patterns are identical to future patterns, so if the patterns are different, the estimated error will be greater and the forecasts will be less accurate.

Korea took an active HIV/AIDS prevention policy with its first diagnosis of HIV infection in 1985, and then expanded HIV testing into specific target groups. Since 1998, Korea has reduced the scope and shifted the policy to voluntary HIV testing with more emphasis on prevention and support of treatment cost of HIV infected persons. This study can be used to estimate the scale of the Korean Government's budget for HIV/AIDS prevention and treatment, as well as inform changes to the HIV/AIDS prevention policy.

Acknowledgements

References

1. UNAIDS report on the global AIDS epidemic 2013. Geneva: UNAIDS; 2013.
2. Korea Center for Disease Control and Prevention. The surveillance reports for HIV/AIDS in Korea 2011. Osong: KCDC; 2013.
3. Reichert TA, Simonsen L, Sharma A, et al. Influenza and the winter increase in mortality in the United States, 1959−1999. Am J Epidemiol 2004 Sep;160(5):492−502.
4. Gaudart J, Toure O, Dessay N, et al. Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali. Malar J 2009 Apr;8:61.
5. Luz PM, Mendes BV, Codeco CT, et al. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. Am J Trop Med Hyg 2008 Dec;79(6):933−9.
6. Yi J, Du CT, Wang RH, et al. [Applications of multiple seasonal autoregressive integrated moving average (ARIMA) model on predictive incidence of tuberculosis]. Chinese Journal of Preventive Medicine 2007 Mar;41(2):118−21 [In Chinese].
7. Liu Q, Liu X, Jiang B, et al. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect Dis 2011 Aug;11:218.
8. Debanne SM, Bielefeld RA, Cauthen GM, et al. Multivariate Markovian modeling of tuberculosis: forecast for the United States. Emerg Infect Dis 2000 Mar−Apr;6(2):148−57.
9. Kee MK, Lee JH, Chu C, et al. Characteristics of HIV seroprevalence of visitors to public health centers under the national HIV surveillance system in Korea: cross sectional study. BMC Public Health 2009 May;9:123.
10. Shin HS, Park JJ, Chu C, et al. Herpes simplex virus type 2 seroprevalence in Korea: rapid increase of HSV-2 seroprevalence in the 30s in the southern part. J Korean Med Sci 2007 Dec;22(6):957−62.
11. Kee MK, Lee JH, Wang J, et al. Ten-year trends in HIV prevalence among visitors to public health centers under the National HIV Surveillance System in Korea, 2000 to 2009. BMC Public Health 2012 Sep;12:831.