

후향연산 모형(Back-calculation model)을 이용한 국내 HIV 감염자와 AIDS 환자의 추계

이주영, 황진수¹⁾, 고운영, 기미경, 김지연²⁾

국립보건원 전염병관리부 역학조사과, 인하대학교 이과대학 수리통계학부¹⁾

Prediction of HIV and AIDS Incidence Using a Back-calculation Model in Korea

Juyoung Lee, Jinsoo Hwang¹⁾, Unyeong Goh, Meekyung Kee, Jeeyun Kim²⁾

Division of Epidemiologic Investigation, Department of Infectious Disease Control, National Institute of Health, Korea;
Department of Mathematics and Statistics, Inha University²⁾

Objective : To estimate the status of HIV infection and AIDS incidence using a back-calculation model in Korea.

Methods : Back-calculation is a method for estimating the past infection rate using AIDS incidence data. The method has been useful for obtaining short-term projections of AIDS incidence and estimating previous HIV prevalence. If the density of the incubation periods is known, together with the AIDS incidence, we can estimate historical HIV infections and forecast AIDS incidence in any time period up to time t . In this paper, we estimated the number of HIV infections and AIDS incidence according to the distribution of various incubation periods

Results : The cumulative numbers of HIV infection from 1991 to 1996 were 708~1,426 in Weibull distribution and 918~1,980 in Gamma

distribution. The projected AIDS incidence in 1997 was 16~25 in Weibull distribution and 13~26 in Gamma distribution.

Conclusions : The estimated cumulative HIV infections from 1991 to 1996 were 1.4~4.0 times more than notified cumulative HIV infections. Additionally, the projected AIDS incidence in 1997 was less than the notified AIDS cases. The reason for this underestimation derives from the very low level of HIV prevalence in Korea. Further research is required for the distribution of the incubation period of HIV infection in Korea, particularly for the effects of combination treatments.

Korean J Prev Med 2002;35(1):65-71

Key Words: Back-calculation, Incubation period distribution, Infection rate, Poisson regression

서론

세계적으로 1999년 12월 말 현재 총 3,430만 명이 인면역결핍바이러스(Human Immunodeficiency Virus : HIV)에 감염된 상태이거나 후천성면역결핍증(Acquired immunodeficiency Syndrome : AIDS)에 이환된 상태로서 추정 감염율은 1.1%이다. 또한 1999년 한 해 동안 540만 명이 새로이 HIV에 감염되어 HIV는 여전히 전세계인의 건강을 위협하는 문제가 되고 있다 [1]. 우리나라의 경우 1985년 국내에서 첫 HIV 감염자가 발견된 이래 2001년 6월말 현재 누적 HIV 감염자수는 1,439명에 이르며, 1997년과 1998년에는 신규감염자가 120명 수준이었으나, 1999년에는 187명, 2000년에는 219명에 이르는 등

증가폭이 커지고 있다 [2].

HIV 감염은 잠복기가 길기 때문에 실제 감염자 규모는 현재 밝혀진 규모보다 많을 것으로 추정되고 있다. 현재의 HIV/AIDS 규모를 파악하고 미래의 감염수준을 예측하는 것은 HIV/AIDS 진단과 치료에 필요한 보건자원을 확보하고 예방정책을 개발하는 등의 관리대책 수립에 필수적인 자료이다. 미국의 경우 HIV 감염자수 추계 연구를 1980년대부터 수행하여 정책수립에 적극적으로 사용하였으며 [3], WHO와 UNAIDS가 주축이 되어 많은 나라에서 HIV/AIDS 규모를 추계하고 예측하는 연구를 활발히 진행해 왔다 [4-6]. 우리나라에서도 1993년~1994년 사이에 세계보건기구 전문가와 국립보건원, 국내 역학전문가 등이 전문가 워크숍을 통하여 HIV 감염

자수를 약 2,500명에서 5,000명 수준으로 추산한 바 있다. 1995년 한국보건사회연구원에서는 약 500명~2,200명으로 추정하였으며 2000년 추계치를 1,460명~6,520명으로 예측한 바 있다 [7].

후향연산 방법은 잠복기가 긴 AIDS 발생자료로부터 과거 감염을 추정하는 방법의 하나이다. 이 방법은 짧은 기간에 AIDS 발생자수를 추계 하거나 HIV 유행률을 추정하는데 효과적이라고 알려져 있다. 후향연산 방법은 Brookmeyer와 Gail [8,9]이 소개하였는데, 본 논문에서는 국내의 AIDS 자료를 바탕으로 후향연산 방법을 이용하여 감염자수를 추계하고, 적절한 잠복기분포를 사용하여 AIDS 발생자수를 예측하였다.

잠복기는 감염된 시점으로부터 AIDS라고 판명된 시점까지의 기간으로 정의한다. 감염되는 시점으로부터 항체양전(seroconversion)되는 기간은 두 달 정도

로 잠복기의 중앙값이 10년인 것에 비하면 매우 짧다고 할 수 있다 [10]. 따라서 본 연구에서는 잠복기를 항체양전 되는 시점으로부터 AIDS로 발전된 시점까지로 정의한다. 잠복기 분포는 외국에서 활용되는 경험적인 잠복기 분포와 Isham [11]에서 언급된 여러 가지 형태의 잠복기 분포를 비교 검토하여 적절한 분포를 이용하여 추계를 실시하였다.

연구대상 및 방법

국내 연도별 AIDS 환자수는 Figure 1과 같다. 본 연구에서는 증가경향을 일관되게 보여준 1991년~1996년 자료를 이용하였다.

1. 잠복기분포

AIDS의 잠복기 분포는 다른 질병에 비하여 매우 길기 때문에 잠복기 분포는 매우 중요하다고 할 수 있다. 본 논문에서 잠복기 분포는 외국에서 활용되는 경험적인 잠복기 분포와 Isham [11]에서 언급된 여러 가지 형태의 잠복기 분포를 사용하여 추계를 실시하였다. 여러 잠복기 분포를 가정하여 환자 수를 추계하고 그 결과를 비교하였다.

잠복기를 추정하는데 있어서 감시체계(surveillance)하에서 수집된 자료가 존재한다고 하면, 그런 자료로부터 항체양전 후 s개월 후의 감염률을 g(s)라고 하자. 또한 이때의 g(s)는 매달 항체양전 후의 변화에 영향을 받지 않는다고 가정하며, s가 0보다 작은 시점에서는 g(s)=0이라고 가정한다.

임의의 확률변수 U가 잠복기간이라고 하고, $F(t)=P[U \leq t]$ 가정하자. 이때, 생존함수 $S(t)=1-F(t)$ 가 된다. 위험률 함수(hazard rate function)는 $\lambda(t)=f(t)/S(t)$ 로 바로 t 이전 시점에서 감염이 없었다는 조건하에 t 시점에서의 AIDS로 발전될 위험률을 나타내는 것이다. 위험률 함수는 감염된 시점으로부터 AIDS로 발전하는 정도를 시간별로 측정하게 된다.

위험률 $\lambda(t)$ 에 대한 모수적인 모형은 역학적 자료를 사용하여서 HIV 감염에 대한 발병에 관련되어 고려되어야 한다. 또한 AIDS로 발전된 후의 위험률 함수는 증가하는 것으로 알려져 있다. 그러나 6년후의 위험률 함수 형태는 AIDS 환자가 치료를 하지 않는 경우가 적기 때문에 위험률 함수의 형태를 알기가 어렵다. 다음은 대표적으로 사용되는 잠복기 분포로 감마분포(Gamma distribution)와

이블분포(Weibull distribution)등을 들 수 있다.

잠복기가 감마분포 $\Gamma(\alpha, \lambda)$ 라고 가정하는 경우에는 다음과 같이 확률밀도함수를 정의할 수 있다.

$$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} \quad \lambda > 0, \alpha > 0$$

이때, $\Gamma(\cdot)$ 은 감마함수이다.

잠복기가 와이블분포 Weibull (p, λ) 라고 가정하는 경우에는 다음과 같이 누적확률밀도함수를 정의할 수 있다.

$$F(t) = 1 - e^{-\lambda t^p} \quad \lambda > 0, p > 0$$

2. 분석방법

1) 후향연산 방법

다음은 Brookmeyer가 소개한 방법이다. 시간 t에 대하여 기대 누적 AIDS 환자수를 A(t)라고 하고, 감염시점 s로부터 감염률을 g(s)이라고 정의하자. 이때 시간 t에 대하여 잠복기 분포를 F(t)라고 하면 다음과 같은 합성곱 방정식(convolution equation)을 정의할 수 있다.

$$A(t) = \int_0^t g(s) \cdot F(t-s) ds \quad (1)$$

위 (1)식은 현재 시점을 t라고 하고, 과거 시점 s에서 감염되었다고 가정하면, 잠복기간이 t-s가 되고, 잠복기 분포 F(t)와 AIDS 환자 발생 누적수 A(t)를 이용하여 감염률 g(s)를 추정하게 된다. 특히, 후향연산 방법의 (1)식 방정식을 사용하기 위해서는 AIDS 누적 환자수 A(t)와 잠복기 분포 F(t)가 주어진 상태에서 추정이 가능하다. (1)식을 적용하여 조사된 환자수 A(t), 잠복기 분포 F(t)를 가지고 감염률 g(s)를 구하고자 한다.

이 과정을 역합성곱(deconvolution)이라 부른다.

j 연도 유행이 시작하는 시점에서 알려지지 않은 감염된 환자의 수를 g_j라 하자. 이때, j 연도에 즉 [j-1, j]기간에 알려진 새로운 AIDS 환자수를 γ_j 라고 하면 각 구간 안에 생겨나는 환자수와 감염자수는 다음 그림과 같이 나타낼 수 있다.

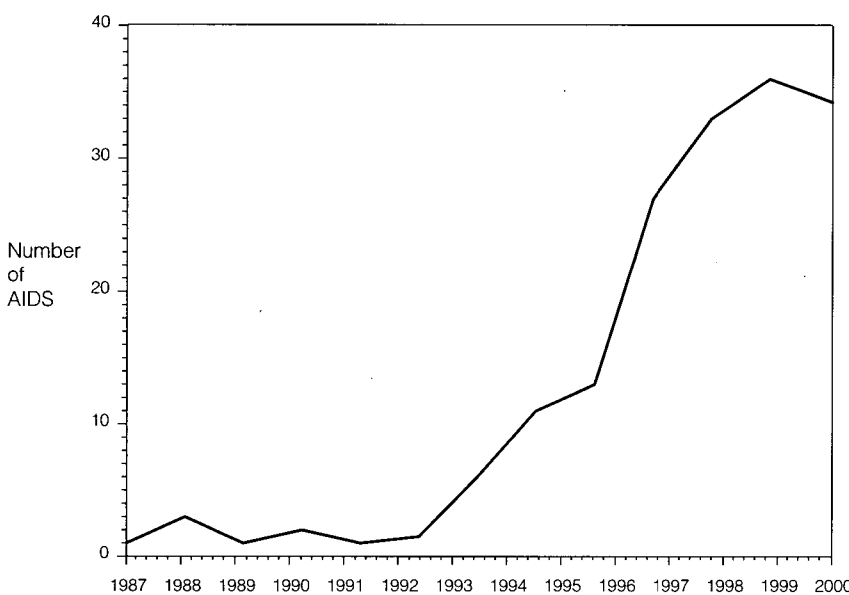
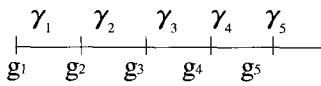


Figure 1. Plot of AIDS incidence in Korea(1987~2000). Source: National Institute of Health, Korea



AIDS 유병률 자료에서 g 들은 추정하고자 하는 감염자수이고 잠복기 분포 $F(t)$ 가 주어지면 그 분포에 따라 선형방정식을 풀게 된다. 감염자가 i 기간에 AIDS환자로 발전하게 될 확률 f_i 을 $f_i=F(i)-F(i-1)$ 라고 하자. 예를 들어, 1년 내에 환자로 발전하게 될 확률은 f_1 이다. 만약 감염자중에서 1년보다 짧은 잠복기를 가지게 되어 환자로 발전하게 된다면 그때의 환자수는 γ_1 이 되게 된다. 따라서, $\gamma_1=g_1 \times f_1$ 이다. 또한 2년 내에 나타난 AIDS 환자수는 잠복기간이 1년 또는 2년 사이이므로 구하고자 하는 AIDS 환자수는 $\gamma_2=g_2f_1+g_1f_2$ 이다. 같은 방법으로 $\gamma_3=g_3f_1+g_2f_2+g_1f_3$ 이고, 따라서, j 년에 AIDS 환자수는 다음과 같이 추정할 수 있다.

$$\gamma_j = g_j f_j + g_{j-1} f_{j-1} + \dots + g_1 f_1 \quad (2)$$

이산일 때의 통계적 역합성곱 방정식은 간단하게 된다. 기대 AIDS 환자수는 다음과 같은 방정식을 만족하게 된다.

$$\mu_j = E(\gamma_j) = g_j f_j + g_{j-1} f_{j-1} + \dots + g_1 f_1 \quad (3)$$

이때, g_j 는 j -년도에 감염자수이고 j 년도에 즉 $[j-1, j]$ 기간에 알려진 새로운 γ_j 는 각 구간 안에 생겨나는 환자수이다.

만일 자료가 정확하게 수집이 되었다면, γ_j 는 포아송(Poisson)분포를 하게 되고, 따라서 γ_j 의 분산은 $E(\gamma_j)$ 와 같아지게 된다. 본 논문에서는 g_j 를 포아송 회귀분석의 회귀계수로 놓고, f_j 는 설명변수로 AIDS 환자수를 종속변수로 보아 포아송 회귀분석을 하여 모형화 하고자 한다. 또한 이모형에서는 절편을 고려하지 않은 모형을 추정하게 되는데 이때의 주의할 점으로는 추정하게 되는 모수가 매우 많아지게 된다는 것이다. 예를 들어, 자료가 n 개가 존재하면 추정하는 회귀계수도 n 개가 되기 때문이다. 모수가 많아지게 되면 추정된 회귀계수의 값이 톱니모양으로 들쭉날쭉하게 될 수도 있고, 불규칙하고 불안정하게 되고, 특히 최근의 감염률에 대한 추정에 대해서는 분산도 커지게 된다 [12].

O' Sullivan [12]은 이러한 현상을 ill-posed라고 정의하였다. 이러한 문제를 해결하기 위해 Brookmeyer [13], Brookmeyer와 Liao [14], Bacchetti와 Segal [15]은 감염률을 매끄러운 곡선이라고 가

정하여 추계 연구를 하였다. 감염률이 매끄러운 곡선(smoothness)이라고 가정하는 경우 앞에서 언급한 문제들을 줄이고, 분산도 줄어드는 것을 보였다. 그러나, 만약 감염률에 대한 가정이 맞지 않다면 추정된 결과가 편향(bias)되어 나타날 수 있다.

본 논문에서 감염률에 대한 가정은 감염률 곡선이 특정한 일부구간에서 특정한 상수라고 가정하는 것이다. 예를 들어, $g_i = g_{i-1} = \beta_i$ 라고 가정 하고, β_i 는 i 년도의 감염률이 된다. $g_i = g_{i-1} = \beta_i$ 의 가정에 의해 처음 3년 동안의 기대 AIDS환자수는 다음과 같이 표현할 수 있다.

$$\begin{aligned} E(\gamma_1) &= \beta_1 f_1 \\ E(\gamma_2) &= \beta_1 f_1 + \beta_2 f_2 \\ &= \beta_1 (f_1 + f_2) \\ E(\gamma_3) &= \beta_1 (f_3 + f_2) + \beta_2 f_3. \end{aligned}$$

일반적으로 $E(\gamma_j)$ 에 대한 표현은 선형이라고 여겨지며 포아송 회귀분석을 이용하여 β 를 추정하게 된다. 위의 가정에 따라 감염곡선을 가정하게 되면 시간에 따른 감염자수는 다음 Figure 2와 같이 나타낼 수 있고, 이것은 2년 간격으로 감염률이 상수라고 가정하는 것이다.

2) 감염률에 대한 모형

후향연산 방법으로 환자수를 추계하는데 있어서 $g(s)$ 의 모수적 모형을 고려하고자 한다. $g(s)$ 의 모수적 모형으로는 크게 강모수 모형(strongly parametric model)과 약모수 모형(weakly parametric model)으로 나눌 수 있다.

강모수 모형(strongly parametric model)으로는 지수적 모형으로서 Taylor [16], De Gruttola와 Lagakos [17]가 사용한 다음과 같은 모형이 있다.

$$g(s) = K\beta_1^{s\beta_2}$$

이모형은 $\beta_2=1$ 인 경우는 지수적으로 증가하고, $\beta_2 < 1$ 이면 지수적으로 증가하는 것보다는 작게 증가한다.

다음으로는 로그-로지스틱모형을 들 수 있는데 이 모형은 다음과 같다.

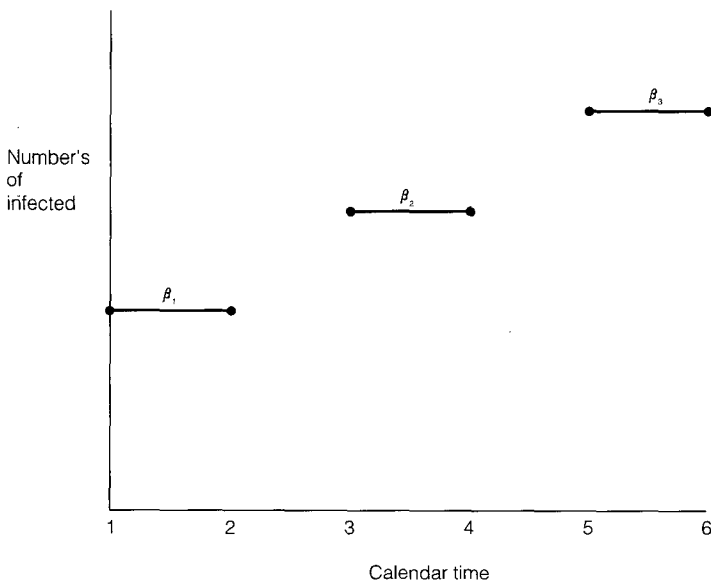


Figure 2. Infection curve in discrete time. Assumption that the numbers of individuals infected, g_j , be constant over specified intervals($g_j = g_{j-1} = \beta_j$).

$$g(s) = K \frac{\beta_1 \beta_2 (s \beta_1)^{\beta_2 - 1}}{(1 + (s \beta_1)^\beta)^2}$$

이모형에서 $\beta_2 > 1$ 이면 감염률이 단조 증가하여 $s = \{(\beta_2 - 1) / (\beta_2 + 1)\}^{1/\beta_1}$ 에서 최고점을 이루게 된다. 만약 $\beta_2 \leq 1$ 이면 감염률은 단조 감소하게 된다. 이모형은 Brookmeyer와 Dimiano [18]가 제안하였다.

강모수 모형에 반하여 약모수 모형 (weakly parametric model)이 존재하는데 예를 들어 계단함수(step-function)같은 구간 안에서는 상수인 그런 함수들이다. 즉 다음과 같은 형태의 함수들이다.

$$g(s) = \beta, \quad c_{i-1} \leq s < c_i$$

이 모형은 감염률이 구간 $[c_{i-1}, c_i)$ 에서 상수 β_i 수준이라고 가정하는 것이다. 따라서, 시간에 따라 구간 안에서는 상수를 이루고 구간이 다음 구간으로 옮겨지면서 감염률이 증가하게 된다. 그러나 이러한 모형은 구간이 달라지면서 감염률의 변화가 크다는 단점이 있다.

또한 이 모형을 고려할 경우 유의해야 할 점은 이러한 변동이 있는 모형을 구간을 어느 정도로 나누어야 하는가 또는 몇 개의 구간으로 나누어야 하는가의 문제이다. 더욱이 최근 감염자에 대한 정보가 부족한 상태이기 때문에 마지막 구간에서 추정된 감염률, β_p 의 표준오차가 구간의 크기가 줄어들수록 커지는 경향이 있다. 반면에 구간의 크기가 너무 크게 되면 실제 감염률이 빠르게 증가하거나 빠르게 감소하는 경우에는 감염률의 편향(bias)이 커지는 현상이 나타나게 된다.

우리는 우리의 AIDS환자자료를 가지고 후향연산 방법을 이용하여 추계함에 있어서 감염률 함수의 모형은 약모수 모형 $g(s) = \beta, c_{i-1} \leq s < c_i$ 를 이용하고 통계적 역합성곱(deconvolution)방법의 포아송 회귀분석 방법으로 추정하였다.

연구결과

먼저 잠복기 분포를 비교하여 보면, 와이블분포와 감마분포의 잠복기간이 큰

Table 1. Median of each incubation period distribution

Incubation distribution	Median (year)	Incubation distribution	Median (year)
Weibull(2.0, 7.3)	5.84	$\Gamma(2, 0.19)$	8.84
Weibull(2.0, 7.5)	6.25	$\Gamma(2, 0.15)$	10.53
Weibull(2.0, 7.6)	6.33	$\Gamma(2, 0.14)$	12
Weibull(1.9, 10)	8.25	$\Gamma(2, 0.12)$	14
Weibull(1.9, 12)	9.9		

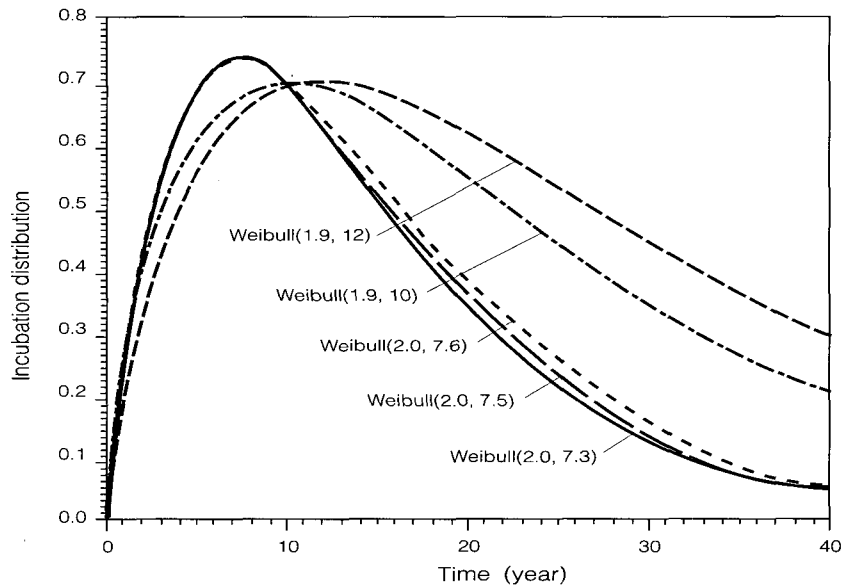


Figure 3. Incubation distribution by Weibull distribution.

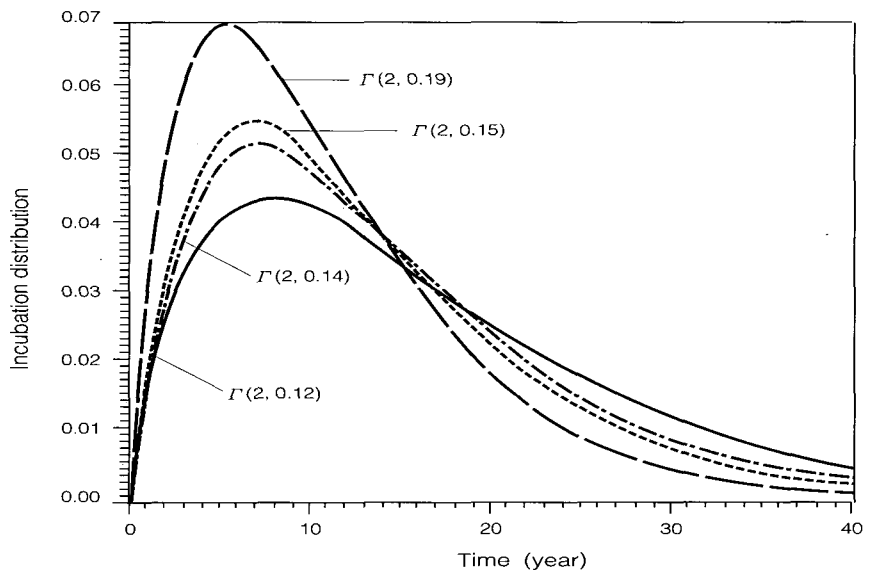


Figure 4. Incubation distribution by Gamma distribution.

차이를 보였다(Table 1).

위 표의 결과를 살펴보면, 와이블분포인 경우 잠복기의 중앙값이 5.84년에서 9.9년이고, 감마분포인 경우 8.84년에서 14년으로 나타나 잠복기간에 대한 중앙값이 감마분포가 와이블분포에 비해 큰 것으로 나타났다.

잠복기간에 대해서 살펴보면 잠복기간이 보통 8년 내지 10년으로 알려져 있으나

의한 잠복기는 더 긴 것으로 나타났다.

따라서, 잠복기간의 중앙값이 9.9년과 10.53년인 Weibull(1.9, 12)와 $\Gamma(2, 0.15)$ 이 잠복기 분포로 가장 좋은 것으로 나타났다. 각 잠복기에 대한 분포는 Figure 3과 Figure 4이다.

본 연구에서는 위에서 언급한 경험적인(국외의 결과) 잠복기 분포와 Isham [11]에서 언급된 여러 가지 형태의 잠복

용된 잠복기분포는 와이블분포와 감마분포로서 대표적으로 이용되는 분포인데 잠복기의 중앙값에 따라서 모수를 다르게 조정하여 다음 결과를 얻었다 (Table 2, Table 3).

잠복기 분포에 따른 추정결과를 살펴보면 와이블분포의 경우나 감마분포의 경우 모두 β_2 의 추정값이 작게 나타나는 것을 알 수 있다. β_2 의 의미는 1993년에서 1994년 사이의 감염자로서 감염자가 줄어드는 현상은 AIDS 자료의 정확도가 낮은 것이 원인이라고 판단된다. 각 잠복기에 의한 1991년~1996년 사이의 신규 HIV 감염자의 합과 1997년의 AIDS환자 수 추계결과는 Table 4와 같다.

위 표의 추정결과는 국내의 수집된 AIDS 환자수에 의해 추정된 결과이다. 6년간의 HIV 감염자의 추계결과는 Weibull(2.0, 7.6)인 경우 $(66+1+313) \times 2=760$ 이다. 또한 7년째의 AIDS 환자 수 추계결과는 $\hat{\gamma}_7 = \beta_1 f_7 + \beta_2 f_7 + \dots + \beta_7 f_7 = 14$ 에 이어 6년째의 감염률이 지속된다고 가정하면 이 가정에 의해 $\beta_7 f_7 = 2$ 이므로 구하고자 하는 다음 7년째의 AIDS환자수는 $14+2=16$ 으로 추계되어진다. 이러한 과정을 반복하여 다른 잠복기에 대한 HIV감염자와 AIDS 환자수를 추계하였다. 1991년부터 1996년까지 국립보건원에 신고된 신규 HIV 감염자의 합은 496명이며 모델에 의한 추정치는 708명~1,980명으로서 신고자수의 1.4~4.0배의 감염자수를 추계하였다. 신고된 1997년 AIDS 환자수는 32명이며 모델에 의해 추계된 1997년 AIDS 환자수는 13~26명으로 보고된 AIDS 환자수보다 적게 추계되었다.

Table 2. Estimated regression coefficients in Weibull distribution

Regression coefficients	Incubation distribution	Weibull	Weibull	Weibull	Weibull	Weibull
		(2.0, 7.3)	(2.0, 7.5)	(2.0, 7.6)	(1.9, 10)	(1.9, 12)
β_1^*		61	55	66	96	135
β_2^{**}		3	2	1	10	6
β_3^{***}		292	306	313	415	572

* β_1 : Number of new HIV infection in 1991~1992

** β_2 : Number of new HIV infection in 1993~1994

*** β_3 : Number of new HIV infection in 1995~1996

국내자료를 추정하여 본 결과 국내 자료에 기 분포를 사용하여 비교하여 보았다. 사

Table 3. Estimated regression coefficients in Gamma distribution

Regression coefficients	Incubation distribution	$\Gamma(2.0, 7.3)$	$\Gamma(2.0, 7.6)$	$\Gamma(1.9, 10)$	$\Gamma(1.9, 12)$
		β_1^*	78	120	136
β_2^{**}	38	38	35	34	
β_3^{***}	343	517	584	772	

* β_1 : Number of new HIV infection in 1991~1992

** β_2 : Number of new HIV infection in 1993~1994

*** β_3 : Number of new HIV infection in 1995~1996

Table 4. Estimated new HIV infections from 1991 to 1996 and estimated new AIDS cases at 1997 by incubation distribution

Incubation distribution	Estimated new HIV infections from 1991 to 1996	New AIDS cases at 1997
Weibull(2.0, 7.3)	708	16
Weibull(2.0, 7.5)	726	16
Weibull(2.0, 7.6)	760	17
Weibull(1.9, 10)	1042	22
Weibull(1.9, 12)	1426	25
$\Gamma(2, 0.19)$	918	13
$\Gamma(2, 0.15)$	1350	26
$\Gamma(2, 0.14)$	1516	18
$\Gamma(2, 0.12)$	1980	21

고찰

1. 결과에 대한 고찰

본 연구에서 이용한 AIDS 환자 추계 결과를 살펴보면 1991년에서 1996년까지 자료를 이용하여 6년간의 HIV 감염자수와 다음 1997년의 AIDS 환자수를 각 잠복기 분포별로 추계하였다. 이 결과인 Table 4를 살펴보면 잠복기를 와이블

분포로 가정한 경우에는 Weibull(1.9, 12)의 경우가 25명으로 가장 근접하며 잠복기의 중앙값도 9.9년인 추정결과를 보이고 있다. 또한 잠복기를 감마분포로 가정한 경우에는 $\Gamma(2.0, 0.15)$ 의 경우가 26명으로 근접한 결과를 보이며 이때 잠복기간의 중앙값은 10.5년으로 나타나 좋은 추정결과를 보이고 있다. 실제 AIDS 잠복기간이 길어지고 있는 국내 상황과 실제 신고된 환자수를 고려할 때 Weibull(1.9, 12)의 경우와 $\Gamma(2.0, 0.15)$ 의 경우가 가장 타당한 잠복기분포로 나타났다. 이러한 결과는 이 자료에 근거한 것이고 더욱 정밀한 추정결과를 얻기 위해서는 지속적인 연구와 정확한 자료의 수집이 요구된다.

HIV 감염자의 결과에서 와이블분포가 708명에서 1,426명사이의 추정값을 갖는 반면 감마분포인 경우는 918명에서 1,980명으로 감마분포로 가정한 경우가 더 큰 값을 갖는 것으로 나타났다. 또한 AIDS 환자수 추계결과도 와이블분포의 경우는 16명이나 17명으로 추정한 결과가 많은 반면 감마분포의 경우에는 13명으로 추정한 결과를 제외하고는 20명 이 넘는 추정결과를 보여 AIDS 환자수 추계결과도 감마분포로 가정한 결과가 더 큰 값으로, 분석한 자료가 다르지만 Isham [11]과 같은 결과로 나타났다.

2. 후향연산 방법에 대한 고찰

Brookmeyer와 Damiano [18]는 CDC(Centers for disease control and prevention)에 신고된 AIDS 환자수 자료를 이용하여 후향연산 방법으로 환자수 추계를 하였다. 이때 사용된 잠복기 분포는 $F(t)=1-\exp(-0.004t^{2.438})$ 이고, Isham [11]이 사용한 잠복기 분포는 $\Gamma(3,0.21)$, $\Gamma(2,0.14)$, Weibull(2,0.12)으로 자료마다 약간의 차이가 있다. Brookmeyer와 Damiano [18]는 각 감염률 모형의 차이에 의한 추계결과를 비교하였고, Isham [11]은 각 잠복기에서의 감염률 곡선이 quadratic exponential 일 경우와 linear logistic 일 경우의 추계결과를 비교하였다.

이러한 후향연산 방법을 사용할 때에

는 몇 가지 주의하여야 할 사항이 있는데 첫째로 미래의 감염률에 대한 정보가 전혀 없고, 단지 과거의 감염률을 추정하는데 적용하는 방법이라는 것이다. 따라서, AIDS 발생자수를 추계 하는데 미래 감염률을 추정하기 위한 좀더 세밀한 가정이 있어야 한다.

만약 미래에 대한 감염률에 대한 정보가 없다면, Brookmeyer와 Gail [8]은 후향연산 방법에 의하여 추정된 AIDS 기대 환자수가 역학조사에서의 최소한의 크기 즉 최소한의 환자수로 볼수 있다고 하였으며, 이에 의해 기대 환자수에 대한 하한(lower-bound)을 얻을 수 있게 된다. 또한, AIDS는 잠복기가 길기 때문에 미래의 감염률이 짧거나 중간정도의 기간에서의 AIDS 환자 수를 추계하는 데는 효과적이지 못하다. 반면에 3~5년전에 또는 더 오래 전에 감염비율 $g(s)$ 를 추계하는 것은 매우 유용하다. 또한 AIDS에 대한 단기간의 추계가 매우 유용하다. 왜냐하면, 추정결과가 최근보다는 좀 떨어진 과거의 감염률에 강하게 영향을 받고 최근 AIDS 유행률로 보정과정을 거쳐서 결과가 얻어지기 때문이다.

두 번째로 최근의 감염률에 대한 정보가 거의 없다는 것이다. 즉, 최근에 진단된 대부분의 AIDS 환자에 대한 정보가 적다는 것이다. AIDS는 잠복기가 긴 질병중의 하나이기 때문에 과거 1년 또는 2년전에 감염된 환자들은 환자수로서 확인할 수 없다는 것이다.

세 번째로, AIDS 잠복기에 대한 분포를 알아야 추정이 가능하다는 것이다. 이 잠복기의 분포에 따라 후향연산 방법의 추정결과는 매우 다양하게 나타나기 때문이다. 더구나 잠복기가 긴 이러한 질병에서는 추정된 감염률이 높게 나타나는데, 수집된 자료가 전체 환자의 아주 작은 일부분으로 나타나기 때문이다. 마지막으로 잠복기에 대하여 잠복기의 분포가 시간에 따라 항상 일정하다는 가정 하에 추정을 하여야 한다는 것이다. 그러나, 실제로 잠복기에 대한 분포가 일정하지 않으며, 가장 중요한 원인은 AIDS 환자를 치료함으로써 환자의 잠복기 분포가 길어

지기 때문이다.

3. 본 연구의 제한점

본 연구에서는 후향연산 방법을 이용하여 다양한 잠복기에 대해 AIDS 환자수를 추계하였다. 이모형은 1987년 이후의 국내자료를 이용해서 얻어진 결과이고, 이 결과는 실제 환자 수보다 과소추정하는 경향을 보였다. 이러한 과소적합의 원인으로는 국내 HIV 감염수준이 매우 낮아 통계모형을 구현하기에 AIDS 환자수가 낮기 때문일 것으로 추정된다. 또한 환자 수를 추계 하는데 있어서 매우 중요한 잠복기분포에 대한 연구가 되지 않아 잠복기에 대한 정보가 정확하지 않았기 때문인 것으로 추정된다. 잠복기가 긴 질병임에 비하여 잠복기에 대한 연구가 빈약한 실정이다. 따라서, 국내 잠복기에 대한 연구가 시급하다고 할 수 있다.

마지막으로 감염률에 대한 모형을 고려함에 있어 $g_i = g_{i+1} = \beta_i$ 을 가정하였다. 환자수의 자료로는 1991년부터 1996년까지 6년간의 자료를 이용하였고 2년 동안 감염률이 동일하다는 가정 하에 모형을 하였으므로 추정할 회귀계수는 $\beta_1, \beta_2, \beta_3$ 3가지가 된다.

회귀분석을 실행함에 있어, 앞에서도 언급한 것과 같이 잠복기분포에 대한 추정이 먼저 진행되어야 하나 잠복기 분포에 대한 연구가 거의 없는 실정이다. 또한 AIDS 환자 수에 대한 추정결과는 잠복기에 따라 그 추정결과가 매우 큰 차이를 보이므로 잠복기에 대한 추정과정이 없는 실정에서 환자 수를 추계하기 어렵다.

이 모형은 감염률이 증가하는 경우에는 그 변화의 추이를 빠르게 따라가지 못하는 단점이 있는 모형이다. 이와 같이 환자 수를 예측하는 경우에는 전년도 감염률을 그대로 사용하면 예년과 양태가 유사하게 진행된다는 가정 하에 구하는 것이므로 만약에 증가추세에 있거나 감소추세에 있다면 이를 적절하게 조절하여야 예측력이 향상될 것이다. Isham [11]은 잠복기의 분포를 감마분포와 와이블 분포로 가정한 경우 감염률 곡선 $g(s)$ 에 대한 모형을 quadratic exponential과

linear logistic으로 가정하여 그 결과를 비교하였는데 국내자료를 이용하여 국내 감염률 모형에 대한 연구가 반드시 필요하다고 할 수 있다.

요약 및 결론

후향연산 방법은 잠복기가 긴 AIDS 발생자료로부터 과거 감염을 추정하는 방법의 하나로서 짧은 기간에 AIDS 발생자수를 추계 하거나 HIV 유병률을 추정하는데 효과적이라고 알려져 있다. 이 방법을 이용하여 국내자료를 가지고 국내 AIDS환자 추계연구를 하였다. 그러나, 국내자료의 부족으로 국내 잠복기에 대한 연구가 거의 없는 실정이므로 외국에서 활용되는 통상적인 경험적인 잠복기 분포와 Isham [11]에서 언급된 여러 가지 형태의 잠복기 분포를 사용하여 비교 검토하여 적절한 분포를 이용하여 추계를 실시하였다.

추계연구를 하는데 1991년부터 1996년까지의 환자자료를 사용하였으며 다양한 잠복기에 대하여 약모수 모형의 감염률을 가정하고 추정을 하였다. 1991년부터 1996년까지 신규 HIV 감염자의 추정치는 708명~1,980명으로서 신고자료의 1.4배~4.0배에 해당되었으며, 1997년

AIDS 환자수는 13~26명으로 1997년 신고 수보다 적게 추계되었다.

참고문헌

- UNAIDS, Report on the global HIV/AIDS epidemic. June. 2000
- 국립보건원. HIV 감염자 발생현황, 2001
- Karon JM, Dondero TJ Jr, Curran JW. The projected incidence of AIDS and estimated prevalence of HIV infection in the United States. *J Acquired Immune Deficiency Syndrome* 1988; 1(6):542-50
- Down AM, Heisterkamp SH, Brunet JB, Hamers FF. Reconstruction and prediction of HIV/AIDS epidemic among adults in the European Union in the low prevalence countries of central and eastern Europe. *AIDS* 1997; 11(5): 649-62
- Commenges D, Etcheverry B. An empirical Bayes approach to the estimation of the incidence curve of HIV infection. *Statistics in Medicine* 1993; 12(14):1317-24
- Schwartzlander B, Stamecki K.A., Brown T., Way P.O., Monasch R. et al. Country-specific estimates and models of HIV and AIDS: methods and limitations. *AIDS* 1999; 13: 2445-2458
- 이순영. 한국의 에이즈 관리 및 감시체계의 효과적 운영방안. 1995, 한국보건사회연구원
- Brookmeyer R, Gail MH. Minimum size of the acquired immunodeficiency syndrome(AIDS) epidemic in the United States. *Lancet* 1986; 2: 1320-22
- Brookmeyer R, Gail MH. A method of obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *JASA*, 1988; 83: 301-08
- Brookmeyer R, Gail MH. AIDS Epidemiology: A Quantitative Approach. Oxford university Press, 1993
- Isham V. Estimation of the incidence of HIV infection, *Philosophical Transaction of the Royal Society of London, Series B*, 1989; 325: 113-21
- O'sullivan F. A Statistical perspective on ill-posed inverse problems. *Statistical Science* 1986; 1: 502-27
- Brookmeyer R. Reconstruction and future trends of the AIDS epidemic in the United States. *Science* 1991; 253: 37-42
- Brookmeyer R, Liao J. Statistical Methods for reconstructing infection curves. In Jewell N, Dietz K, Farewell V(eds), *AIDS Epidemiology: Methodologic Issues*, 1992, pp 39-60. Boston: Birkhäuser
- Bacchetti P, Segal M, Jewell NP. Back-calculation of HIV infection rates(with discussion), *Statistical Science*, 1993
- Taylor JM. Models for the HIV infection and AIDS epidemic in the United States, *Statistics in Medicine* 1989; 8: 45-58
- De Gruttola V, Lagakos SW. Analysis of doubly-censored survival data with applications to AIDS, *Biometrics* 1989a; 45: 1-11
- Brookmeyer R, Damiano A. Statistical methods for short-term projections of AIDS incidence, *Statistics in Medicine* 1989; 8: 23-34