

후향연산식을 활용한 국내 삼일열 말라리아의 감염분포와 유병자수 추정

장현갑¹ · 박정수² · 전미정³ · 이정애⁴ · 김한메울⁵

¹원자력병원; ²전남대학교 통계학과; ³대한적십자사 광주전남혈액원;
⁴전남대학교 의과대학 예방의학과; ⁵전남대학교 통계학과

(2008년 7월 접수, 2008년 8월 채택)

요약

국내 삼일열 말라리아는 1990년대 중반부터 급격히 증가하여 2000년대에는 연평균 1600여명이 발병하고 있다. 본 연구에서는 이미 알려진 잠복기 분포와 2001년부터 2006년까지의 발병자수 자료에 바탕하여 후향연산식을 활용하여 국내 삼일열 말라리아의 감염분포를 최소제곱법으로 추정하였다. 추정된 감염분포는 평균이 207일이고 표준편차가 30.7일인 정규분포를 이루었다. 이를 이용하여 연간 유병자수 분포를 산출한 결과, 말라리아의 하루 평균 유병자수는 628.8명이었다.

주요어: 몬테카를로 기법, 발병자수 분포, 생존함수, 잠복기 분포, 정규분포, 최소제곱법.

1. 서론

세계적으로 약 30억 명이 말라리아 감염 위험에 노출되어 있고 매해 3억 5천만 명에서 5억 명이 말라리아에 감염되며 백만 명 이상이 사망한다. 말라리아는 21세기에도 여전히 전 세계의 큰 숙제이다 (WHO, 2005). 국내에는 1910년까지 말라리아가 널리 유행하였고, 신의학의 도입과 함께 점차 감소되어 1984년 이후에 근절된 것으로 여겨졌다. 그러나 1993년 과주지역의 군사보호구역(DMZ)에 복무중인 군인 1명이 말라리아 환자로 발생했고, 이후 계속 증가하여 2000년에 4,142명으로 정점을 보인 뒤 2004년 826명까지 감소 추세를 이어가다가 2005년과 2006년에는 각각 1,369명과 1,997명의 환자가 발생해 다시 증가 추세를 보이고 있다 (그림 1.1). 국내 삼일열 말라리아(*Plasmodium vivax malaria*)는 지역적으로 대개 군사보호구역을 따라 발생하고 있다 (Lee 등, 2002; Park 등, 2003). 그러나 말라리아 토착지역을 여행한 사람에서도 말라리아가 발병되고 있기 때문에 비말라리아 토착지역 거주자에게도 역시 위협이 되고 있다 (Kim, 2001).

국내 삼일열 말라리아는 매개체인 중국얼룩날개모기(*Anopheles sinensis*)의 계절적 개체수 변동과 겨울철 동면에 따라 명확한 계절성을 보인다. 또한 3개월 이하의 단기 잠복기와 특징적인 1년 가까이의 장기 잠복기를 가진다 (Nishiura 등, 2007). 국외에서 발병되는 열대열(*Plasmodium Falciparum*) 말라리아

¹(139-706) 서울특별시 노원구 공릉동 215-4, 원자력병원, 의사. E-mail: nicholas_10@hanmail.net

²교신저자: (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 교수. E-mail: jspark@chonnam.ac.kr

³(530-370) 광주광역시 남구 송하동 127-4, 대한적십자사 광주전남혈액원 의무관리실장, 의사.

⁴(501-746) 광주광역시 동구 학1동 5, 전남대학교 의과대학 예방의학과, 교수. E-mail: jarhee@healthis.org

⁵(500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 석사과정. E-mail: meury@hanmail.net

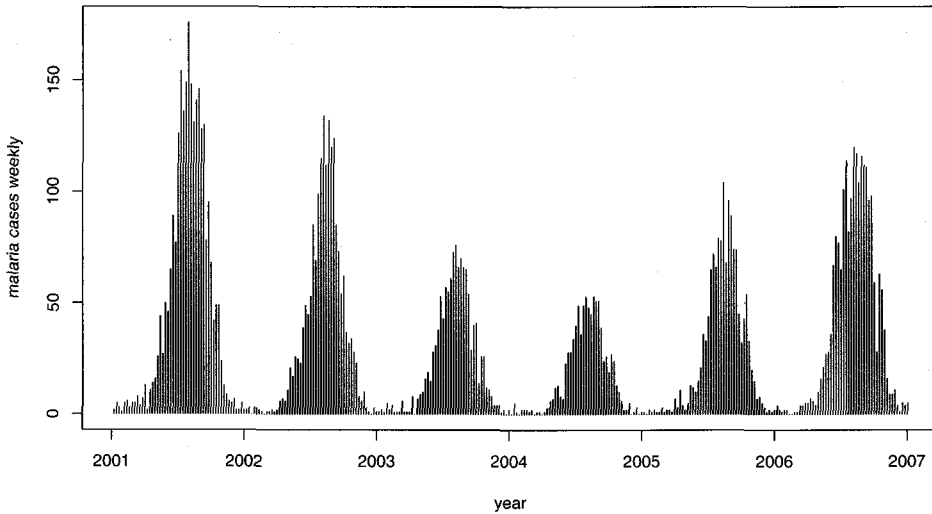


그림 1.1. 국내 삼일열 말라리아의 6년간(2001년-2006년) 주별 발병자수 시계열

이는 거의 대부분 2개월 내에 발병하여 단기 잠복기만을 가진다. 따라서 국내 말라리아가 실제로 감염 되는 시기의 분포는 질병관리본부에서 조사한 발병자수 분포와는 다른 형태일 것으로 추측된다.

본 연구에서는 질병관리본부의 2001년부터 2006년까지의 발병자수 자료(질병관리본부 전염병 웹통계)와 Nishiura 등 (2007)이 제시한 잠복기 분포를 이용하여, 몬테카를로 기법으로 감염분포의 대략적인 형태(정규분포)를 파악하고, 후향연산식을 활용하여 최소제곱법으로 정규분포의 모수를 추정했다. 이렇게 추정된 감염분포와 생존함수를 이용하여 유병자수를 추정했다. 여기서 감염분포란 1월 1일부터 12월 31일 사이에 감염자수의 분포(distribution of infected patients)를 말한다.

본 논문은 장현갑 (2008)의 석사학위 논문에 기초하고 있지만 내용과 통계적 방법에서 훨씬 개선된 것이다. 이정희 등 (2008)은 국내 신증후군 출혈열 자료에 대해 본 논문과 비슷한 방법으로 감염분포와 유병자수를 추정했다. 신증후군 출혈열은 잠복기가 짧고 단순하여 본 연구보다 간단한 방법이 이용됐다. 후향연산기법은 대부분 AIDS 환자 자료 (Day 등, 1989; Becker 등, 1991; Brookmeyer, 1991)나 약물 주사 관련 자료 (De Angelis 등, 2004)에 대해 연단위로 적용되었다. 이주영 등 (2002)은 발병자수가 포아송 분포를 따른다고 보고 모수적 감염률 함수와 후향연산식에 바탕한 포아송 회귀분석을 실시하여 국내 HIV 감염자와 AIDS 환자를 추정했다. Bacchetti 등 (1993)은 HIV 감염률을 추정하는데 감염이 비균등 포아송 과정을 따른다고 보고 거침에 대한 벌칙 우도함수(roughness-penalized likelihood)를 최초로 하는 추정법을 사용했다.

2. 국내 삼일열 말라리아의 감염분포 추정

2.1. 발병자수 분포

질병관리본부에서 매일 말라리아 발병자수를 집계하기 시작한 2001년부터 2006년까지의 자료를 이용했다(질병관리본부 홈페이지). 국외 감염자를 제외한 국내 감염에 의한, 6년간 총 발병자수는 9817명(연평균 1637명)이다. 발병시기의 원자료는 요일과 공휴일에 따라서 편차가 매우 크므로, 원자료 대신

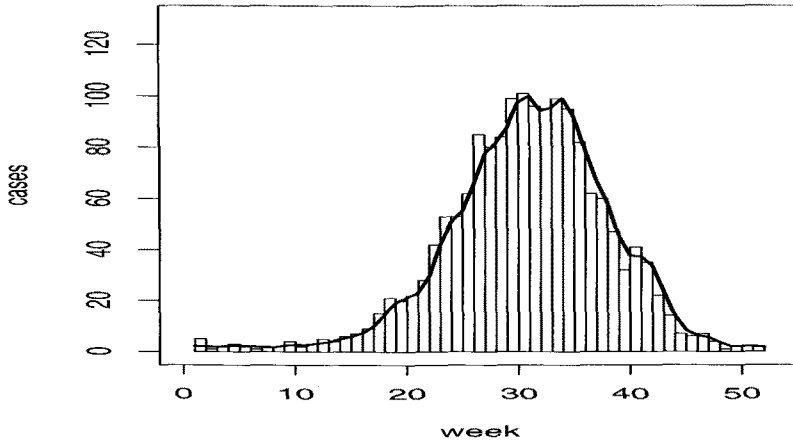


그림 2.1. 국내 삼일열 말라리아의 과거 6년간의 평균 발병자수에 대한 연중 주별 막대그림과 15일 이동평균선

15일 이동평균을 사용했다. 그림 2.1은 과거 6년간의 평균 발병자수에 대한 연중 주별 막대그림과 이동평균선이다. 그림은 정점을 향하여 왼쪽에서 서서히 올라갔다가 정점에서 오른쪽으로 비교적 급하게 내려오는 경향을 보인다. 평균은 214.4일(30.6주)이고 표준편차는 50.6일(7.23주)이다. 대략 15주째(4월 중순)부터 발병이 증가하여 44주째(11월 초순)에 급감한다.

2.2. 잠복기 분포

Nishiura 등 (2007)은 한국 삼일열 말라리아의 잠복기 분포를 구했다. 이들의 연구는 국내의 2000년-2003년에 말라리아 비위험지역에서 위험지역으로 여행한 사람들을 대상으로 하여 여행시기와 발병시기를 조사하여 잠복기를 구했다. 확률변수 U 를 잠복기간이라 할 때 U 는 다음과 같이 63.1%의 감마분포와 36.9%의 정규분포의 혼합분포를 따랐다 (그림 2.2):

$$f(u) = .631 \text{ Gamma}(1.2, 22.2) + .369 N(337.4, 40.6^2). \tag{2.1}$$

단기 잠복기의 평균은 26.64일이고 표준편차는 24.32일이다. 잠복기간 U 에 대해 $F(t) = \text{Pr}(U \leq t)$ 라고 할 때, $S(t) = 1 - F(t)$ 는 생존함수이다. 즉 말라리아에 감염되고 나서 t 일 동안(발병하지 않고) 병을 가지고 있을 확률이다.

2.3. 발병자수의 후향연산식

발병자수 분포와 잠복기 분포로부터 감염분포를 추정해내는 잘 알려진 방법은 후향연산기법(back-calculation method)이다 (Brookmeyer와 Gail, 1986; 1988). 후향연산기법에서 감염분포를 구하는 계산은 역행렬을 이용하는데, 이 경우 구해진 감염분포가 매우 들쭉날쭉하고 음수를 갖는 경우도 있다 (장현갑, 2008; O’Sullivan, 1986). 이러한 문제를 해결하기 위해 Brookmeyer와 Liao (1992), Bacchetti 등 (1993)은 감염분포가 매끄러운 곡선이라고 가정했다. 한편 감염에 대해 어떤 모수적 분포를 가정하고 후향연산식에서 얻은 예측치와 실제 발병자수와 차이의 합이 최소가 되도록 모수를 정해주는 최소제곱법도 생각할 수 있다. 본 연구는 후자를 택했는데, 어떤 분포를 사용할 것인가를 결정하기 위해 몬테카를로 기법을 활용했다(자세한 기술은 부록에 제시함).

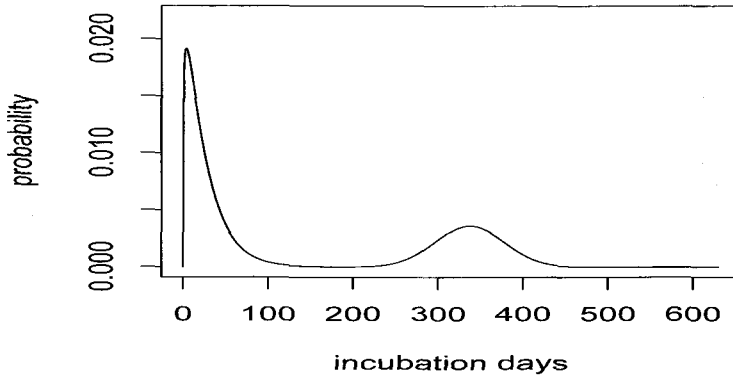


그림 2.2. NISHIURA 등 (2007)에 의해 구해진 국내 삼일열 말라리아의 잠복기 분포

부록에서 감염분포가 정규분포를 따른다는 근거를 보였고, 정규분포의 모수를 추정하는데 후향연산식을 활용한다. 1년 중의 날 $d = 1, 2, \dots, 365$ 에 대하여, A_d^y 를 y 년도 d 일의 발병자수, G_d^y 를 y 년도 d 일의 감염자수, f_d 를 식 (2.1)로 주어진 d 날간의 잠복 확률이라고 하자. 잠복기의 최대기일이 730일이라는 가정 하에서 A_d^y 는 $y - 2$ 년도 d 날 이후에 감염된 사람들 중에서 y 년도 d 날에 발병한 사람수이다. 어느 고정된 날 d 에 대하여, A_d^y 를 구하는 후향연산식을 유도하면 다음과 같다.

먼저 $y - 2$ 년도 d 날부터 $y - 2$ 년도 365일 사이에 감염된 사람들 중에 y 년도 d 날에 발병한 사람수는

$$f_{730} G_d^{y-2} + f_{730-1} G_{d+1}^{y-2} + \dots + f_{730-(365-d)} G_{365}^{y-2}$$

이다. 또 $y - 1$ 년도 첫 날부터 $y - 1$ 년도 마지막일 사이에 감염된 사람들 중에 y 년도 d 에 발병한 환자수는

$$f_{365+d-1} G_1^{y-1} + f_{365+d-2} G_2^{y-1} + \dots + f_{365+(d-365)} G_{365}^{y-1}$$

이다. 그리고 y 년도 첫 날부터 y 년도 $d - 1$ 일 사이에 감염된 사람들 중에 y 년도 d 날에 발병한 사람수는

$$f_{d-1} G_1^y + f_{d-2} G_2^y + \dots + f_{d-(d-1)} G_{d-1}^y$$

이다. 이때

$$G_d^y = n^y g_d \tag{2.2}$$

인데, n^y 는 y 년도의 총 감염자수이고 g_d 는 d 날에 감염될 확률이다. 가정 $g_{730+d} = g_{365+d} = g_d$ 와 식 (2.2)을 이용하고, 위의 세 식을 합하면 다음과 같다.

$$A_d^y = \sum_{k=0}^{365-d} f_{730-k} n^{y-2} g_{d+k} + \sum_{k=365-d+1}^{730-d} f_{730-k} n^{y-1} g_{d+k} + \sum_{k=730-d+1}^{729} f_{730-k} n^y g_{d+k}, \tag{2.3}$$

여기서는 $n^{y-2} = n^{y-1} = n^y \equiv n$ 이라고 가정하므로, 식 (2.3)는

$$A_d^y = n \sum_{k=0}^{729} f_{730-k} g_{d+k} \tag{2.4}$$

가 된다. 식 (2.4)의 양변을 모든 년도와 날에 대해 더하면

$$\begin{aligned} T &= \sum_{y=1}^Y \sum_{d=1}^{365} A_d^y = Yn \sum_{k=0}^{729} f_{730-k} \left(\sum_{d=1}^{365} g_{d+k} \right) \\ &= Yn \sum_{k=0}^{729} f_{730-k} = Yn \end{aligned}$$

이므로 $n = T/Y$ 이다. 본 연구의 경우 $Y = 6$ 이고 $T = 9817$ 이다. 따라서 $B_d = \sum_{y=1}^Y A_d^y$ 라고 표기하면 식 (2.4)로부터, $d = 1, 2, \dots, 365$ 에 대하여

$$B_d = Yn \sum_{k=0}^{729} f_{730-k} g_{d+k} \tag{2.5}$$

가 된다. B_d 는 실제 자료로부터 과거 Y 년간 d 날의 총 발병자수로 추정되고 이 추정량을 \hat{B}_d 으로 표기한다.

2.4. 최소제곱법에 의한 감염분포의 모수 추정

모든 \hat{B}_d 와 f_{730-k} 가 주어졌으므로 이를 이용하여 감염확률 g_d 를 구한다. 감염분포가 정규분포 모양을 띤다는 것을 알므로 모수들을 최소제곱법으로 구했다. 즉

$$\sum_{d=1}^{365} \left\{ \hat{B}_d - T \sum_{k=0}^{729} f_{730-k} g_{d+k}(\mu, \sigma) \right\}^2 \tag{2.6}$$

을 최소로 하는 μ 와 σ 를 구했다. $g_{d+k}(\mu, \sigma)$ 는 $N(\mu, \sigma^2)$ 의 확률밀도함수를 말한다.

실제 자료(15일 이동평균을 취한 발병자수 자료)에 대해 의사-뉴턴(quasi-Newton) 최소화 알고리즘을 적용하여 구한 결과, $\hat{\mu} = 207.3$ 일(29.6주), $\hat{\sigma} = 30.7$ 일(4.4주)이다. 이 감염분포는 발병자수 분포에 비하여 분산이 작고, 또한 몬테카를로 기법으로 구한 감염분포보다도 훨씬 분산이 작으며 최빈값에서의 높이가 높았다. 감염의 평균이 발병의 평균보다 약 7일 정도 빠르게 나타났다. 감염은 20주 또는 21주째(5월 하순)에서 시작되어 38주 또는 39주째(9월 하순)까지 이어진다. 이는 말라리아 모기의 활동성으로 보아 합당한 분포로 여겨진다.

3. 국내 삼일열 말라리아의 유병자수 추정

P_d^y 를 y 년도 d 일의 유병자수, G_d^y 를 y 년도 d 일의 감염자수, S_t 를 잠복기간 U 의 생존함수라고 하자. 즉 S_t 는 t 일간 유병할 확률이다. 잠복기의 최대일일이 730일이라는 가정 하에서, P_d^y 는 $y - 2$ 년도 d 일 이후에 감염된 사람들 중에서 y 년도 d 일까지 유병하는 사람수이다. 어느 고정된 날 d 에 대하여, P_d^y 를 구하는 후향연산식을 유도하면 다음과 같다.

먼저 $y - 2$ 년도 d 날부터 $y - 2$ 년도 365일 사이에 감염된 사람들 중에 y 년도 d 일에도 유병인 사람수는

$$G_d^{y-2} S_{730} + G_{d+1}^{y-2} S_{730-1} + \dots + G_{365}^{y-2} S_{730-(365-d)}$$

이다. 또 $y - 1$ 년도에 감염되어 y 년도 d 일에도 유병인 사람수는

$$G_1^{y-1} S_{365+d-1} + G_2^{y-1} S_{365+d-2} + \dots + G_{365}^{y-1} S_{365+(d-365)}$$

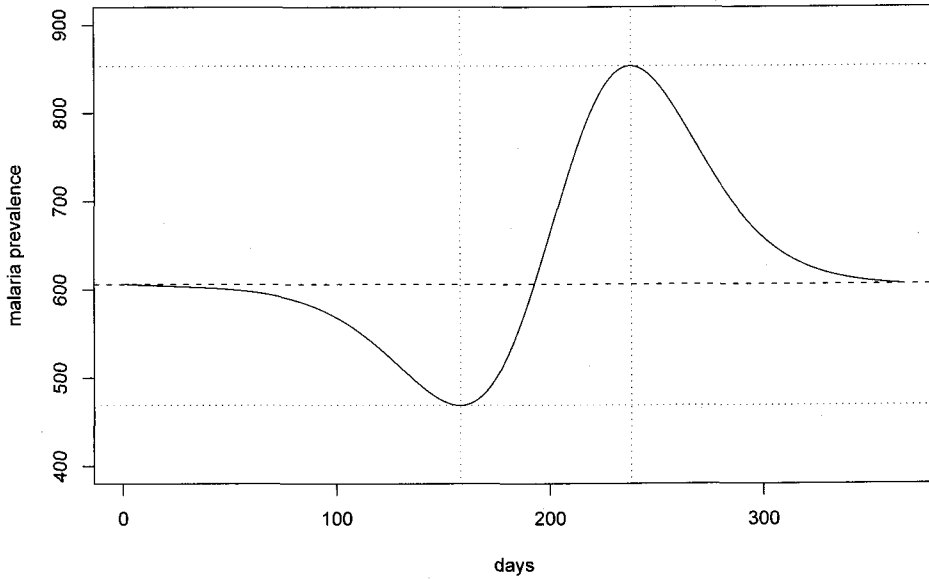


그림 3.1. 국내 삼일열 말라리아의 연중 유병자수 분포(158일(22.6주)째에 최저값 469명, 193일(27.6주)째에 겨울의 유병자수(606명)로 되었다가 238일(34주)째에 최고값 853.5명을 가지는 비대칭 분포)

이다. 그리고 y 년도에(d 일 이전에) 감염되어 d 일에도 유병인 사람수는

$$G_1^y S_{d-1} + G_2^y S_{d-2} + \dots + G_{d-1}^y S_{d-(d-1)}$$

이다. 이 세가지를 합하면 다음과 같다.

$$P_d^y = \sum_{t=d}^{365} G_t^{y-2} S_{730-(t-d)} + \sum_{t=1}^{365} G_t^{y-1} S_{365+d-t} + \sum_{t=1}^{d-1} G_t^y S_{d-t}. \quad (3.1)$$

본 연구에서는 $G_t^{y-2} = G_t^{y-1} = G_t^y \equiv G_t$ 를 가정하므로 식 (3.1)은 $d = 1, 2, \dots, 365$ 에 대하여

$$P_d = \sum_{t=d}^{365} G_t S_{730-(t-d)} + \sum_{t=1}^{365} G_t S_{365+d-t} + \sum_{t=1}^{d-1} G_t S_{d-t} \quad (3.2)$$

이 된다. 이때 식 (2.2)로부터 $G_t = ng_t$ 이다.

생존함수와 앞 절에서 추정한 감염분포를 이용하여 구한 유병자수 분포는 그림 3.1이다. 하루 평균 유병자수는 628.8명, 표준편차는 103.4명이고, 158일(22.6주)째에 최저값 469명을, 238일(34주)째에 최고값 853.5명을 가진다. 연중 분포는 겨울에 일정하게(약 606명으로) 유지되다가 봄에 내려가고 감염이 시작되는 초여름에 급격히 높아졌다가 가을에 다시 낮아지는 형태를 띤다. 또한 193일(27.6주)째에 겨울의 유병자수로 되었다가 그 뒤로 더욱 높아지는 비대칭을 보인다.

그림 3.1을 자세히 해석하면, 겨울 유병자수는 장기 잠복기에 있는 사람수로 대략 연 평균 발병자수 1637×0.369 에서 나온 숫자이다. 이들이 초봄부터 서서히 발병하여(장기 잠복기인 정규분포의 평균이 337일로 365일 보다 작음에 유의) 치료되다가 6, 7월에 새로 감염되는(곧바로 발병하지는 않는) 사람들 때문에 빠르게 높아지지만 늦가을로 갈수록 감염자가 서서히 줄어들고 그 사이 63.1% 정도는 발병

하여 치료가 되므로 유병자수도 감소하는 양상을 보였다. 연중 유병자수 분포가 27.6주째(6월 중순)를 중심으로 비대칭을 이루는 이유는 그 때 이후로 새로히 감염되는 비율이(감염분포의 평균인) 29.6주에서 절정을 이루고, 게다가 평균 3.8주의 단기 잠복기를 거치기 때문이다. 즉 $29.6\text{주} + 3.8\text{주} = 33.4\text{주}$ 로 유병자수가 최고로 되는 34주째와 가깝게 된다.

4. 결론과 토의

국내 삼일열 말라리아는 1990년대 중반부터 급격히 증가하여 2000년대에는 연평균 1600여명이 발병하고 있다. 국내 말라리아는 잠복기의 복잡성(3개월 이하의 단기 잠복기와 1년 가까이 장기 잠복기의 혼합)때문에 국외 말라리아에 비하여 감염분포의 추정이나 유병자수 계산이 상대적으로 복잡하다. 본 연구에서는 이미 알려진 국내 삼일열 말라리아의 잠복기 분포 (Nishiura 등, 2007)와 발병자수 자료에 바탕하여 몬테카를로 기법을 이용하여 감염분포를 추정했으나 적절하지 못하여 정규분포에 가깝다는 것만을 먼저 알았다 (부록). 후향연산식을 활용하여 최소제곱법으로 말라리아의 감염확률인 정규분포의 모수를 추정했다. 그 결과, $\hat{\mu} = 207.3\text{일}(29.6\text{주})$, $\hat{\sigma} = 30.7\text{일}(4.4\text{주})$ 이었다. 감염은 20주 또는 21주째(5월 하순)에서 시작되어 38주 또는 39주째(9월 하순)까지 이어지는 형태를 취했다.

또한 생존함수와 감염분포를 이용하여 유병자수 분포를 구했다 (그림 3.1). 하루 평균 유병자수는 628.8명, 표준편차는 103.4명이고, 158일(22.6주)째에 최저값 469명을, 238일(34주)째에 최고값 853.5명을 보였다. 연중 분포는 겨울에 일정하게(약 606명으로) 유지되다가 봄에 내려가고 감염이 시작되는 초여름에 급격히 높아졌다가 가을에 다시 낮아지는 형태를 띠었다. 또한 193일(27.6주)째에 겨울의 유병자수로 되었다가 그 뒤로 더욱 늘어나는 비대칭을 보였다. 이러한 감염양상과 유병자수 분포를 활용하는 적절한 예방 사업이 필요하다고 생각한다.

부록에서 몬테카를로 기법을 이용하여 감염분포가 대략 정규분포를 따른다는 근거를 보인 것은 무작정 정규분포를 가정하는 것보다 바람직하다. 또한 감염분포를 구하기 위한 단순한 몬테카를로 기법은 부분적으로 옳지 않은 결과를 낸다는 것을 보여줬다. 이에 대해 한 심사자는 식 (A.2)로부터 감염분포가 대략 꼬리가 약간 두꺼운 정규분포를 취할 것이 당연하다고 언급했지만 이를 수리적으로 증명하기는 쉽지 않아 보인다. 근거도 없이 어떤 분포를 가정하거나 수리적 증명도 없이 정규분포가 당연하다고 말하는 것보다는 연구자가 가능한 방법을 동원하여 어떤 근거를 보이는 것이 연구에서 더 필요하다. 한편 부록의 그림 A.1을 보고 혹자는 정규분포보다는 감마분포가 더 적절하다고 할 수 있다. 실제로 장현갑(2008)은 역행렬 계산의 결과로부터 감염분포가 대략 감마분포를 따른다고 보았다.

모수를 추정하는데 최소제곱법 대신에 최소 카이제곱 추정법도 가능한 방법일 것이다. 장현갑(2008)은 감염분포와 유병자수 추정에 바탕하여, 현재 적십자에서 수혈전과성 말라리아를 막기 위해 실시하는 현혈 문진과 효소면역법(ELISA)의 비용-효과 분석 방법을 제시했다. 한편 Jang 등(2008)은 국내 말라리아 자료에 대하여 본 논문의 방법과는 다르게, 포아송 모형과 최우추정법을 이용하여 감염분포를 추정했다. 또한 장기 잠복기 환자의 경우 발현되기 전까지는 현혈 등에 의한 전염이 되지 않는다는 가설에 바탕한 전염가능 유병자수(prevalence with parasitemia)를 계산하는 방법을 새롭게 유도하고 실제 자료에 적용했다.

부록. 몬테카를로 기법으로 구한 감염분포

감염일에 잠복기를 더하면 발병일이 되는데, 이미 발병일과 잠복기 분포가 주어졌으므로 감염일은

$$\text{감염일} = \text{발병일} - \text{잠복기} \quad (\text{A.1})$$

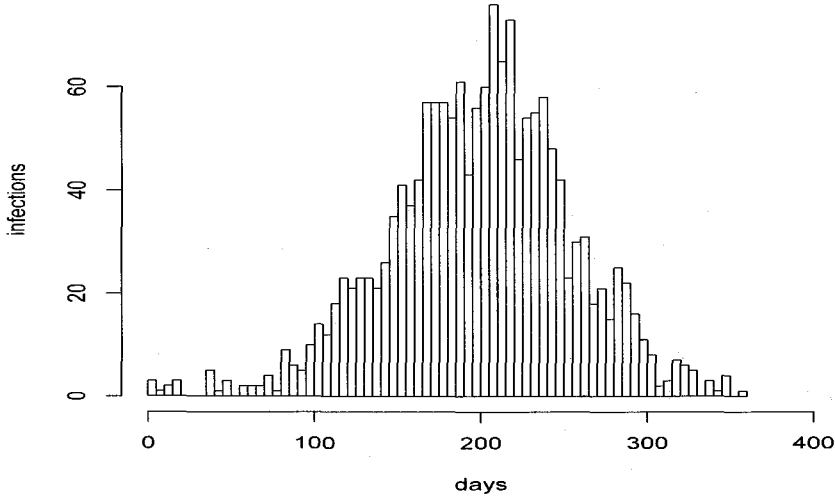


그림 A.1. 몬테카를로 기법으로 구한 감염분포(상식적으로 맞지 않게 분산이 확대되었으나 대략적인 분포 형태를 알 수 있다)

라는 공식을 이용하여 구했다. 이를 위한 몬테카를로 알고리즘은 다음과 같다.

[단계 1] 2.1절에서처럼 발병자수(과거 6년간의 평균) 원자료로부터 j 날의 15일 이동평균 발병자수 a_j ($j = 1, 2, \dots, 365$)를 구한다.

[단계 2] 하나의 j 에 대하여, 잠복기 분포 (2.1)로부터 난수를 발생시켜서 그것을 C_j 라고 하자($1 \leq C_j \leq 730$). 감염일은 식 (A.1)로부터

$$I = [j - C_j + 730] \pmod{365} \quad (\text{A.2})$$

으로 구한다. 하나의 j 에 대하여 난수 C_j 를 발생시키고 I 를 구하는 작업을 a_j 번 반복 실행한다 (여기서 I 들은 j 날에 발병한 환자들의 감염일들이다.).

[단계 3] 위 [단계 2]를 모든 $j = 1, 2, \dots, 365$ 에 대해 실행하여 얻어진 I 들에 대하여, $G_j = \#\{I \mid I = j\}$ 를 구한다. G_j 는 j 날 감염자수이다.

[단계 2]의 식 (A.2)에서 $+730 \pmod{365}$ 를 한 이유는 $j - C_j$ 값이 -729 부터 0 사이 값을 가질 수 있기 때문에 I 가 1 부터 365 사이 값을 갖도록 하기 위함이다. 이때 730 일(2년)까지 고려한 이유는 장기 잠복기가 1년을 넘어서기 때문이다.

위 알고리즘을 적용하여 구한 감염분포 그림은 대략 정규분포 모양을 보였다 (그림 A.1 참조). 그러나 그림 A.1은 말라리아 모기가 활동하지 않는 겨울과 초봄, 늦가을에도 상당수(1년에 약 160여명)가 감염 되는 현상을 나타냈는데, 이는 상식적으로 맞지 않다.

이 결과는 위와 같은 단순한 몬테카를로 기법으로는 감염분포를 잘 유도해내기 곤란함을 보여준다. 다만 감염분포가 대략 정규분포를 따른다는 근거만을 알아냈다. 여기서 몬테카를로 기법을 이용하여 감염분포가 대략 정규분포를 따른다는 근거를 보인 것은 무작정 정규분포를 가정하는 것보다 바람직하다. (저자들은 본 연구의 시작에서 위와 같은 몬테카를로 기법으로 감염분포를 추정할 수 있으리라 예상했으나 상식적으로 맞지 않게 분산이 팽창된 결과를 얻게 되어, 결국 후향연산식을 활용하게 됐다.)

참고문헌

- 이정희, 이영생, 박정수 (2008). 국내 신증후군 출혈열의 감염자 분포와 유병자수 추정, *Journal of the Korean Data Analysis Society*, **10**, 1915-1922.
- 이주영, 황진수, 고운영, 기미경, 김지연 (2002). 후향연산 모형(Back-calculation model)을 이용한 국내 HIV 감염자와 AIDS 환자의 추계, <예방의학회지>, **35**, 65-71.
- 장현갑 (2008). 감염확률밀도함수를 이용한 말라리아 유병률 추정, <전남대학교 대학원 보건학 협동과정 석사학위 논문>, 광주.
- 질병관리본부, 홈페이지: <http://stat.cdc.go.kr>
- Bacchetti, P., Segal, M. R. and Jewell, N. P. (1993). Backcalculation of HIV infection rate, *Statistical Science*, **8**, 82-119.
- Becker, N. G., Watson, L. F. and Carlin, J. B. (1991). A method of non-parametric back-projection and its application to AIDS data, *Statistics in Medicine*, **10**, 1527-1542.
- Brookmeyer, R. (1991). Reconstruction and future trends of the AIDS epidemic in the United States, *Science*, **253**, 37-42.
- Brookmeyer, R. and Gail, M. H. (1986). Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States, *Lancet*, **2**, 1320-1322.
- Brookmeyer, R. and Gail, M. H. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic, *Journal of the American Statistical Association*, **83**, 301-308.
- Brookmeyer, R. and Liao, J. (1992). Statistical methods for reconstructing infection curves, *AIDS Epidemiology: Methodologic Issues*, Jewell N., Dietz K., Farewell V.(eds), 36-60, Birkhauser, Boston.
- Day, N. E., Gore, S. M., McGee, M. A. and South, M. (1989). Predictions of the AIDS epidemic in the U.K.: The use of the back projection method, *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **325**, 123-134.
- De Angelis, D., Hickman, M. and Yang, S. (2004). Estimating long-term trends in the incidence and prevalence of opiate use/injecting drug use and the number of former users: Back-calculation methods and opiate overdose deaths, *American Journal of Epidemiology*, **160**, 994-1004.
- Jang, H. G., Lee, J. A. and Park, J. S. (2008). Estimating infection distribution and prevalence with parasitemia of *Plasmodium vivax* malaria using incidence and incubation period, Technical report, Department of Statistics, Chonnam National University, Gwangju.
- Kim, M. B. (2001). Epidemiologic characteristics of malaria in non-malarious area, Jeollabuk-do, Korea in 2000, *The Korean Journal of Parasitology*, **39**, 223-226.
- Lee, J. S., Lee, W. J., Cho, S. H., and Ree, H. I. (2002). Outbreak of vivax malaria in areas adjacent to the demilitarized zone, South Korea, 1998, *The American Journal of Tropical Medicine and Hygiene*, **66**, 13-17.
- Nishiura, H., Lee, H. W., Cho, S. H., Lee, W. G., In, T. S., Moon, S. U., Gyung, T. C. and Kim, T. S. (2007). Estimates of short- and long-term incubation periods of *Plasmodium vivax* malaria in the Republic of Korea, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **101**, 338-343.
- O'Sullivan, F. A. (1986). A statistical perspective on ill-posed inverse problems, *Statistical Science*, **1**, 502-527.
- Park, J. W., Klein, T. A., Lee, H. C., Pach, L. A., Ryu, S. H., Yeom, J.S., Moon, S. H., Kim, T. S., Chal, J. Y. Oh, M. D. and Choe, K. W. (2003). Vivax malaria: A continuing health threat to the Republic of Korea, *The American Journal of Tropical Medicine and Hygiene*, **69**, 159-167.
- WHO (2005). *World Malaria Report*.

Estimating Infection Distribution and Prevalence of Malaria in South Korea Using a Back-calculation Formula

Hyun Gap Jang¹ · Jeong-Soo Park² · Mi Jeong Jun³ · Jeong Ae Rhee⁴ · Han Meury Kim⁵

¹Dept. of Education and Training, Korea Cancer Center Hospital;

²Dept. of Statistics, Chonnam National University;

³Korea Red Cross, Gwangju-Chonnam Blood-Bank;

⁴Dept. of Preventive Medicine, Chonnam National University;

⁵Dept. of Statistics, Chonnam National University

(Received July 2008; accepted August 2008)

Abstract

Incidence of *Plasmodium vivax* malaria in South Korea have been reemerged from mid-1990 and infected around 1600 patients annually recent years. The authors calculated the distribution of malaria infection and prevalence in South Korea using incidence (2001–2006) and incubation period distributions by a back-calculation formula and the least squares estimation method. The estimated infection has a normal distribution with a mean 207 and a standard deviation 30.7 days. In addition, the authors found the estimated daily average prevalence is 628.8 patients.

Keywords: Incidence distribution, incubation period distribution, least squares estimation, Monte-Carlo technique, normal distribution, survival function.

¹Doctor, Dept. of Education and Training, Korea Cancer Center Hospital, 215-4 Gongneung-dong, Nowon-gu, Seoul 139-706, Korea. E-mail: hg.jang@naver.com

²Corresponding author: Professor, Dept. of Statistics, Chonnam National University, 300 Yongbong-dong, Bukgu, Gwangju 500-757, Korea. E-mail: jspark@chonnam.ac.kr

³Doctor, Korea Red Cross, Gwangju-Chonnam Blood-Bank, Gwangju, Korea.

⁴Professor, Dept. of Preventive Medicine, College of Medicine, Chonnam National University, Gwangju 530-370, Korea. E-mail: jarhee@healthis.org

⁵Graduate Student, Dept. of Statistics, Chonnam National University, Gwangju 500-757, Korea. E-mail: meury@hanmail.net