

제2장

생명표 사망확률 추정방법 개선연구 - 시도 생명표와 사망원인 생명표를 중심으로 -

백지선 · 정미옥

제1절 서론

1. 연구배경 및 목적

생명표(Life table)는 남녀별 연령에 따른 사망자수, 생잔률, 기대여명, 평균수명 등을 통계적 방법으로 작성하여 주어진 인구집단의 사망 상태를 보여주는 지표들로 작성된 통계표이다(구자홍, 2009). 우리나라는 1960년 인구센서스를 계기로 1964년 최초로 생명표를 작성하였으며, 1990년부터 인구동태 신고자료를 기초자료로 하여 생명표를 작성해 오고 있다. 생명표는 전국인구를 대상으로 하는 일반생명표(또는 국민생명표)가 있으며, 사망원인 생명표, 혼인생명표와 같이 특정한 일부 인구집단을 대상으로 하는 특수생명표가 있다. 생명표는 작성방법에 따라 각 세별로 작성되는 완전생명표(Complete Life Table)와 보통 5세 간격으로 작성되는 간이생명표(Abridged Life Table)로 구분할 수가 있다. 현재 통계청에서는 일반생명표 외에 시도 생명표, 사망원인 생명표를 작성하여 제공하고 있다. 시도 생명표는 16개 시도에 대해 해당 지역의 사망수준을 보여주도록 작성되며, 사망원인 생명표는 특정사인에 의한 사망확률과 특정사인이 제거된 경우 증가하는 연령별 기대여명 등을 나타낸다. 생명표의 작성과정은 (1)관찰기간 설정과 그에 따른 연령별 사망자수와 인구수 등의 기초자료 수집 및 자료 구성 (2)기초자료를 이용한 연령별 기초 사망률 산출 (3)이를 근거로 하여 보정된 사망확률 산출 (4)주어진 사망질서(사망확률)와 작성기법에 따른 생명표 확정 단계로 구성된다. 마지막 생명표 확정단계에서는 10



만명의 0세 인구를 시작으로 폐쇄인구 가정하에 일련의 공식들을 이용하여 생명표를 완성하는 것이므로, 그 전 단계인 관찰기간 설정 및 기초 자료 구성과 사망확률 산출은 생명표 작성의 기초이며, 중요한 작업이라고 할 수 있다.

관찰기간은 현재의 상황을 잘 설명할 수 있는 기간으로 설정해야 하는데 매년 작성되는 일반생명표의 경우는 관찰기간을 1년으로 해도 충분하다. 그러나 세분화된 지역이나 인종 등에 대해 작성하는 경우는 1년이라는 기간이 부족할 수가 있으며 이러한 경우에는 보다 긴 관찰기간이 필요할 것이다. 사망확률의 보정(graduation)은 주어진 인구집단에 대한 사망확률을 추정할 때 정확성(fitness)을 유지하면서 평활성(smoothness)을 높여주는 통계적 기법을 이용하여 사망질서를 일반적이고 안정적인 형태로 나타내고자 하는 것이다. 사망확률 보정방법은 여러 가지가 있으며, 우리나라는 Greville 9차항 방법을, 미국은 Beers 방법을 이용하고 있다. 박유성 외(2009)는 Greville 보정방법과 Beers 보정방법을 우리나라 자료에 적용하여 정확성과 평활성 측면에서 비교하였는데, 정확성 부분에서는 Greville 9차항 보정식이 우수하며 평활성 부분에서는 Beers 보정식이 우수함을 보였다.

우리나라 시도 생명표는 관찰기간을 1년으로 하여 일반생명표와 동일한 방법으로 완전생명표와 간이생명표 형태로 작성되고 있으며, 미국과 캐나다의 경우 관찰기간 3년에 대해 지역을 구분하여 완전생명표와 간이생명표를 작성하고 있다. 현재 우리나라 시도 생명표는 기초자료가 충분치 못하여 각 세 사망확률이 불안정하게 산출될 수 있는 가능성을 가지고 있으므로 안정적인 사망확률이 산출될 수 있도록 관찰기간과 각 세 사망확률 산출방법에 대한 연구가 필요하다.

사망원인 생명표는 일반적으로 특정사인에 대한 사망확률표와 특정사인 제거시 생명표라는 두 가지 형태로 작성된다. 특정사인에 대한 사망확률표는 특정사인으로 인해 그 연령층의 사람들이 궁극적으로 사망하게 될 확률(또는 사인으로 인한 사망 위험률)을 나타낸다. 특정사인 제거시 생명표는 특정사인이 제거되었다는 가정 하에서 다른 사인으로 인해 언젠가는 사망하게 되는 경우에 대해 작성되는 것으로 특정사인이 제거됨으로써 신장되는 국민의 기대여명을 제공한다. 이러한 사망원인 생명표의 의미에 맞추어 현재 작성되고 있는 방법과 이론적 배경을 검토하여 타당성을 확인하고 재정립하는 것이 필요하다. 본 연구에서는 현재 작성되고 있는 시도 생명표와 사망원인 생명표의 사망확률 산출 및 작성방법을 체계적으로 검토하여 보다 안정적이고 타당한 사망확률 추정방안을 마련하고자 한다.

2. 연구내용

시도 생명표와 사망원인 생명표는 각 시도와 사망원인에 대한 사망구조를 나타내며, 지역과 의료정책 수립에 필수적인 기본정보를 제공하므로 정확하고 안정적인 결과를 도출하도록 작성되어야 한다. 우리나라 시도 생명표는 작성하고자 하는 해에 해당하는 단년도 실측치에 Greville 9차항 보정을 통해 산출된 사망확률을 이용하여 2005년부터 3년 주기로 작성되고 있다. 안정적인 시도 생명표의 사망확률 산출방법 검토를 위하여 관찰기간을 1년(2008년)과 3년(2007~2009년)으로 설정하고 보정방법을 Beers 보정방법과 Greville 9차항 보정에 비해 평활도가 높은 13차항 보정을 적용하여 현재의 생명표 결과 및 정확도와 평활도 측면에서 비교하였다. 사망원인 생명표는 특정 사망원인에 대한 위험률과 특정사인 제거시 사망확률에 대한 근거와 작성방법을 이론적으로 정리하였으며, 실증분석으로 2008년 남자에 대해 사인이 악성신생물인 경우의 사망원인 생명표를 작성하였다.

보고서 구성은 다음과 같다. 2절에서는 통계청의 시도별, 사망원인별 작성방법을 살펴보았다. 3절에서 주요 나라에 대한 시도별, 사망원인 생명표 작성주기 및 방법을 검토하였다. 4절에서는 검토한 방법들을 중심으로 시도 생명표와 사망원인 생명표의 개선에 대해 논의하고 우리나라 자료에 적용한 후 현재의 방법과 비교하였다. 5절에서는 연구 결과를 정리하였다.

제2절 통계청 생명표

본 절에서는 현재 통계청에서 작성되고 있는 연간 전국단위의 각 세 일반생명표와 시도단위로 작성되는 시도 생명표 그리고 주요 사인별로 작성되는 사망원인 생명표의 작성방법에 대하여 살펴보고자 한다.

1. 생명표 작성방법과 용어

생명표는 주어진 인구에 대해 산출된 연령별 사망확률을 이용하여 폐쇄인구를 가정하고 10만명의 0세 인구(기수인구)를 시작으로 주어진 사망구조(연령별 사망확률)에 따라 인구가 소멸되어가는 과정을 나타내는 통계표이다. 연령별 사망확률은 각 세 혹은 5세 간격으로 작성된다. 통계청(2010)을 참고하여 생명표 작성과 용어를 간략하게 정리하



면 다음과 같으며, n 은 연령구간을 나타낸다.

(사망확률) ${}_nq_x = x$ 세인 사람이 $(x+n)$ 세에 이르지 못하고 사망할 확률

(생존자수) $l_x = x$ 세에 생존할 것으로 기대되는 사람 수 ($l_0 = 100,000$ 명)

$$= l_{x-1} - {}_nd_{x-1}$$

(사망자수) ${}_nd_x = x$ 세에 생존한 사람이 $(x+n)$ 세에 이르지 못하고 사망할 것으로 기대되는 사람 수 $= l_x \times {}_nq_x$

(생존확률) ${}_nS_x = x$ 세인 사람이 $(x+n)$ 세까지 생존할 확률 $= l_{x+n}/l_x$

(정지인구) ${}_nL_x = x$ 세의 생존자들이 $(x+n)$ 세에 도달하는 n 년 기간 동안 살아줄 것으로 기대되는 생존년수

* 정지인구는 사망사건이 균등하게 발생된다는 가정 하에 산출되며, 0세의 경우 이러한 가정을 적용할 수 없기 때문에 0세 사망분리계수 f_0 를 이용하여 산출함

$${}_nL_x = (l_x + l_{x+1})/2$$

$${}_1L_0 = f_0 \times l_0 + (1 - f_0) \times l_1$$

(총생존년수) $T_x = x$ 세의 생존자들이 모두 사망할 때까지 살 것으로 기대되는 생존

$$\text{년수의 합계} = \sum_x^w {}_nL_x, \quad w = \text{최고연령}$$

(기대여명) $e_x = x$ 세인 사람이 향후 생존할 것으로 기대되는 평균 생존년수

$$= T_x / l_x$$

(기대수명) e_0 는 0세 출생자의 평균 생존년수로 0세의 기대여명

2. 일반생명표와 시도 생명표

우리나라 일반생명표(국민생명표)는 1980년에 인구동태표본조사 결과를 이용하여 1978~1979년 생명표를 작성한 것을 시작으로 현재는 1년을 작성주기로 매년 완전생명표와 간이생명표를 작성하여 제공하고 있다. 시도 생명표는 16개 시도에 대해 2005년에 처음으로 작성되었으며 3년의 작성주기로 2008년에 이어 2011년에 작성된다. 생명표 작성을 위한 기초자료는 일반생명표의 경우 전국 사망자수와 주민등록인구이며, 시도 생명표는 시도별 사망자수와 주민등록인구이다. 일반생명표와 시도 생명표는 기초자료만 차이가 있고 작성방법은 동일하므로, 구분하지 않고 소개한다. 생명표 작성에 필요한 연령별 사망확률은 기초자료를 보정한 후 사망률과 사망확률을 계산하고 Greville 9차항 보정식을 적용하여 산출되며, 자세한 방법은 아래의 <표 2-1>에 수록하였다.

<표 2-1> 생명표의 연령별 사망확률표

(0세 사망확률)

- 영아사망자수(d) : 성별 영아 사망자수
- 영아사망확률(q_0) : 0세 사망 분리계수(f_0 :영아사망자중 전년도 출생아 비율)를 산출하여 영아사망확률로 전환

$$q_0 = d \times \frac{1-f_0}{b^t} + d \times \frac{f_0}{b^{t-1}}$$

b^t : 당해년도 출생아수, b^{t-1} : 전년도 출생아수,

$$f_0 = \frac{t\text{년도}0\text{세사망자중}t-1\text{년도출생아수}}{t\text{년도}0\text{세사망자수}}$$

(1세~99세 사망확률)

- 연령별 사망률을 사망확률로 전환 : $q'_x = \frac{m_x}{1 + \frac{1}{2}m_x}$



- 그레빌 9차항 계수를 이용하여 연령별 사망확률을 보정

$$q_x = -0.040724q'_{x-4} - 0.009873q'_{x-3} + 0.118470q'_{x-2} + 0.266557q'_{x-1} \\ + 0.331140q'_x + 0.266557q'_{x+1} + 0.118470q'_{x+2} \\ - 0.009873q'_{x+3} - 0.040724q'_{x+4}, \quad (x = 1, 2, \dots, 99)$$

$$q'_x = 1.352613q'_{x+1} + 0.114696q'_{x+2} - 0.287231q'_{x+3} - 0.180078q'_{x+4} \\ (x = 0, -1, -2, -3)$$

(85세이상 사망확률의 보정)

- 고연령에서는 사망확률 증가세가 둔화된다는 경향을 이용하여

$${}_nq_x = {}_nq_{x-1} \times e^{k_x} \text{의 관계식 이용하여 추정}$$

- k_x 는 연령(x)과 k 값의 관계를 선형관계($k_x = ax + b$)로 가정

이와 같이 산출된 사망확률이 기초가 되어 기수인구 100,000명을 시작으로 일반적인 생명표 작성방법에 따라 상한 연령을 100세로 설정하여 0~100+세 완전생명표를 작성한 후 간이생명표를 작성한다. 5세(간이)생명표의 연령별 생존자수(l_x), 총생존년수(T_x), 연령별 기대여명(e_x) 값은 완전생명표의 해당연령(x)의 값과 각각 동일하며 연령구간별 사망자수, 정지인구, 사망확률은 각각 아래의 관계식을 적용하여 산출한다.

$${}_nd_x = \sum d_x, \quad d_x: \text{완전생명표의 연령별 사망자수}$$

$${}_nL_x = \sum L_x, \quad L_x: \text{완전생명표의 연령별 정지인구}$$

$${}_nq_x = {}_nd_x / l_x, \quad l_x: \text{완전생명표의 연령별 생존자수}$$

시도 생명표의 경우 상한 연령을 85세 이상으로 조정하여 0,1,5,~,85+세 시도별 간이 생명표를 작성한 후 시도별 사망원인 생명표를 작성하게 된다.

3. 사망원인 생명표

사망원인 생명표는 1995년부터 매년 간이생명표(0,1,5,10~85+세) 형태로 작성되고 있다. 사망원인 생명표는 특정사인에 대한 사망자수, 특정사인 제거 사망자수, 특정사인 제거 생존자수, 특정사인 제거 정지인구, 특정사인 사망확률, 특정사인 제거 기대여명 및 특정사인 제거 증가 기대여명으로 구성된다. 여기서 특정사인 사망확률은 연령별 특정사인에 의해 궁극적으로 사망할 확률로 위험률을 의미한다. 사망원인 생명표 작성을 위한 기초자료는 사망원인을 구분하지 않은 간이생명표와 5세 단위 사망자수, 사망원인 별 사망자수이며, 작성방법과 용어를 정리하면 다음과 같다.

(특정사인에 의한 사망자수)

${}_n d_x(i)$ = 연령구간 $(x, x+n)$ 세 사이의 사망자 중 특정사인 i 에 의한 사망자수

$${}_n d_x(i) = {}_n d_x \times ({}_n D_x(i) / {}_n D_x),$$

여기서 ${}_n D_x$ = x 세와 $x+n$ 세 사이에 관측된 사망자수

${}_n D_x(i)$ = x 세와 $x+n$ 세 사이에 관측된 사망자 중 사인 i 에 의한 사망자수

(특정사인 사망확률)

$R_x(i)$ = x 세인 사람이 궁극적으로 특정사인으로 사망하게 될 확률

$$R_x(i) = \sum_{t=x}^w \frac{{}_n d_t(i)}{l_x}, \quad \omega \text{는 상한 연령}$$

(특정사인 제거시 사망자수)

${}_n d_x(-i)$ = 연령구간 $(x, x+n)$ 세 사이의 사망자 중 특정사인 i 에 의한 사망자수를 제외한 사망자수

$${}_n d_x(-i) = {}_n d_x - {}_n d_x(i)$$

(특정사인 제거시 생존자수, 생존년수, 총생존년수, 기대여명)

$l_x(-i)$, ${}_nL_x(-i)$, $T_x(-i)$, $e_x(-i)$ = 일반생명표와 동일한 개념으로, 특정사인에 의한 사망자수를 제외한 사망자수에 의해 작성

$$l_{x+n}(-i) = l_x(-i) - {}_nd_x(-i)$$

$${}_nL_x(-i) = n \times \left(\frac{l_x(-i) + l_{x+n}(-i)}{2} \right), \quad {}_\infty L_{85}(-i) = \frac{l_{85}(-i)e_{85}}{1 - R_{85}(i)}$$

$$T_x(-i) = \sum_x^w {}_nL_x$$

$$e_x(-i) = T_x(-i)/l_x(-i)$$

(특정사인 제거시 증가기대 여명)

$e_x(-i) - e_x$ = 특정사인으로 인해 감소되는 기대수명

(참고) 0세와 1~4세의 정지인구는 사망사건이 균등하게 발생된다는 가정을 적용할 수 없으므로 인해 분리계수(f_0 , f_{1-4})를 산출하여 다음과 같이 산출한다.

$$(0\text{세}) \quad {}_1L_0(-i) = f_0 \times l_0(-i) + (1 - f_0) \times l_1(-i)$$

$$(1\text{-}4\text{세}) \quad {}_4L_1(-i) = f_{1-4} \times l_1(-i) + (4 - f_{1-4}) \times l_5(-i),$$

$$f_{1-4} = \sum_1^4 (x - 0.5) \times d_x / \sum_1^4 d_x$$



제3절 해외 사례연구

생명표 작성방법에 대하여 미국과 캐나다의 경우 비교적 자세하게 제공하고 있으므로, 본 절은 미국과 캐나다의 생명표 작성방법을 중심으로 소개하기로 한다. <표 2-2>에서 <표 2-4>까지 미국, 캐나다, 일본¹⁾의 생명표 작성 주기 및 자료에 대해 정리 수록하였다.

〈표 2-2〉 전국생명표 비교

구분	미국	캐나다	일본
생명표 종류	① Decennial ② Annual	Complete	① Complete ② Abridged
작성주기	① 10년 ② 1년	5년	① 5년 ② 1년
나이구분 (최근)	①② 각 세(0~100+)	각 세(0~109)	① 각 세(0~open ended) ② 각 세(0~105+)
영유아 구분	①② 1, 7, 28, 365일	1, ..., 7일, 2, 3, 4주, 2, ..., 12개월	①② 2, 3, 6개월
이용자료	①② 인구 센서스 인구동태통계 의료보험자료	인구 센서스 인구동태통계	① 인구동태통계 인구 센서스 ② 잠정인구동태통계 인구 추정치
관찰기간	① 3년 ② 1년	3년	①② 1년

1) 출처 : <http://www.mhlw.go.jp/>



〈표 2-3〉 지역별 생명표 비교

구분	미국	캐나다	일본
지역수준	주(state)	① 주(Province) ② 준주(Territories)	① 도도부현(Prefecture), ② 시정촌(Municipality)
작성주기	10년	5년	5년
나이구분 (최근)	각 세(0~100+) *영아는 1, 7, 28, 365일	① 각 세(0~109) *지역별 자료품질에 따라 105~109사이에서 절단 ② 5세(0,1-4,~,95-99,100+)	
이용자료	인구 센서스 인구동태통계 의료보험자료	인구 센서스 인구동태통계	
관찰기간	3년	3년	
제외지역	주별 인종별 성별 사망자수 700명 이하	Prince Edward Island는 인구와 사망자수가 적어 간이생명표로 작성됨	

〈표 2-4〉 사망원인 생명표 비교

구분	미국	캐나다	일본
작성주기	10년	1년	1년
나이구분 (최근)	5세(0-1, ..., 100+)	5세(0-1, ..., 90+)	0세, 65세, 75세, 90세
이용자료	완전생명표의 간이버전 사망원인별 사망자수		
관찰기간	3년	1년	1년
보고서/ 공표내용		Mortality, Summary List of Causes/ 지역별 # of death, mortality rate per 100,000 population	Abridged Life Table/ mortality probability, potential years of life lost(a certain cause of death is eliminated)

1. 미국 생명표

미국의 경우 연간 완전생명표를 해당 해의 자료를 이용하여 전국수준에서 작성하고 있다. 또한 이와는 별도로 센서스 해를 중심으로 그 전후에 해당되는 3년 자료를 이용하여 10년 주기의 완전생명표를 전국과 주(state) 수준에서 작성하고 있다. 사망원인 생명표도 10년 주기로 3년간의 자료를 이용하여 작성하고 있다. 먼저 연간 완전생명표 작성법에 대해 NCHS(1999)를 참고하여 1996년도 사례를 정리하였다.

가. 연간 완전생명표

미국의 연간 완전생명표의 상한 연령은 100세 이상이며, 성별(남/여)과 인종별(백인/흑인)로 생명표가 작성되고 있다. 생명표는 사망확률(q_x)로부터 계산되는데, 사망확률은 해당 해에 관측된 사망자수(D_x)와 연앙인구(P_x)에 기초하여 구해진다. 기초자료 준비단계와 사망확률을 계산하는 방법을 검토하고, 일련의 생명표 작성방법을 정리해 보았다.

1) 기초자료

사망확률을 추정하기 위해 각 세별 인구동태통계와 센서스의 관측값을 그대로 사용하게 되면 사망시 연령보고 착오와 같은 이례적인 상황에 의한 문제가 발생할 수 있다. 이러한 예외적인 상황을 제거하기 위해 일반적으로 보정을 통해 연령별 사망확률 곡선을 평활화한다. 미국에서는 Beers 방법(Beers ordinary minimized fifth difference formula)에 의해 5세 간격의 인구자료를 아래와 같은 공식을 사용하여 각 세별로 내삽(interpolation)하였다.

$$P_{x+k} = \begin{cases} C_{k,0} {}_5P_0 + C_{k,5} {}_5P_5 + C_{k,10} {}_5P_{10} + C_{k,15} {}_5P_{15} + C_{k,20} {}_5P_{20}, & x = 0, 5 \\ \begin{aligned} &C_{k,x-10} {}_5P_{x-10} + C_{k,x-5} {}_5P_{x-5} + C_{k,x} {}_5P_x \\ &+ C_{k,x+5} {}_5P_{x+5} + C_{k,x+10} {}_5P_{x+10} \end{aligned}, & x = 10, \dots, 90 \\ \begin{aligned} &C_{k,80} {}_5P_{80} + C_{k,85} {}_5P_{85} + C_{k,90} {}_5P_{90} + C_{k,95} {}_5P_{95} \\ &+ C_{k,100} {}_5P_{100} \end{aligned}, & x = 95 \end{cases}$$

여기서 $k = 1, 2, 3, 4$ 이고, P_{x+k} 는 $x+k$ 세의 인구수이며, ${}_5P_x$ 는 x 세에서 $x+5$ 세 사이의 인구수이다. $C_{k,x}$ 는 Beers 계수로 <표 2-5>와 같다.

〈표 2-5〉 Beers 계수

연령		해당 연령으로 시작하는 5세 그룹				
0~9세		0	5	10	15	20
0	+0.3333	-0.1636	-0.0210	+0.0796	-0.0283
1	+0.2595	-0.0780	+0.0130	+0.0100	-0.0045
2	+0.1924	+0.0064	+0.0184	-0.0256	+0.0084
3	+0.1329	+0.0844	+0.0054	-0.0356	+0.0129
4	+0.0819	+0.1508	-0.0158	-0.0284	+0.0115
5	+0.0404	+0.2000	-0.0344	-0.0128	+0.0068
6	+0.0093	+0.2268	-0.0402	+0.0028	+0.0013
7	-0.0108	+0.2272	-0.0248	+0.0112	-0.0028
8	-0.0198	+0.1992	+0.0172	+0.0072	-0.0038
9	-0.0191	+0.1468	+0.0822	-0.0084	-0.0015
5x~5x+5세		x-10	x-5	x	x+5	x+10
x	-0.0117	+0.0804	+0.1570	-0.0284	+0.0027
x+1	-0.0020	+0.0160	+0.2200	-0.0400	+0.0060
x+2	+0.0050	-0.0280	+0.2460	-0.0280	+0.0050
x+3	+0.0060	-0.0400	+0.2200	+0.0160	-0.0020
x+4	+0.0027	-0.0284	+0.1570	+0.0804	-0.0117
95~99세		80	85	90	95	100
95	-0.0015	-0.0084	+0.0822	+0.1468	-0.0191
96	-0.0038	+0.0072	+0.0172	+0.1992	-0.0198
97	-0.0028	+0.0112	-0.0248	+0.2272	-0.0108
98	+0.0013	+0.0028	-0.0402	+0.2268	+0.0093
99	+0.0068	-0.0128	-0.0344	+0.2000	+0.0404

사망자수 자료의 경우 먼저 전체 사망자수(D) 중 연령미상 사망자수(D^a)의 비율인 조정인자($F = D/D^a$)를 계산하여 연령그룹별로 사망자수에 조정인자를 곱해줌으로써 연령미상 사망자수에 대한 보정을 하였다. 이와 같이 연령미상이 보정된 5세 간격의 사망자수를 인구자료와 마찬가지로 Beers 공식에 의해 각 세로 내삽하였다. 단, 5~9세, 10~14세에 대해서는 다른 연령대와는 다르게 ${}_5D_0$ 대신에 가공(fictitious)의 값을 이용하



여 내삽하였다. 이는 실제 관측된 ${}_5D_0$ 를 사용한 5~14세의 내삽 값과 0~4세의 사망자수의 연결이 매끄럽지 못할 수 있기 때문에 ${}_5D_0$ 에 대한 가공의 값(${}_5D_0^*$)을 아래와 같이 계산하여 사망자 수 대신 사용하며, V 는 2~4세 사망자수의 합이다.

$${}_5D_0^* = 2.45580 V - 0.59332 {}_5D_5 - 0.01965 {}_5D_{10} + 0.22004 {}_5D_{15} - 0.08055 {}_5D_{20}$$

2) 사망확률

사망확률은 0세, 1~84세, 85~99세 그룹별로 각기 다른 방식으로 구하고 있다. 0세 사망확률은 출산 코호트 방법에 의해 아래와 같이 계산하였다.

$$q_0 = \frac{D_0(1-f)}{B^t} + \frac{D_0f}{B^{t-1}}$$

여기서 f 는 분리계수로 t 해에 사망한 영아 중 $(t-1)$ 해에 출생한 영아의 비율이며, D_0 는 영아 사망자수, B_t 는 t 해의 출생자 수이다.

1~84세 사망확률은 아래의 식에 의해 계산하였다. 1~4세의 경우 D_x 는 연령미상을 보정한 관측된 사망자수이고, P_x 는 Beers 공식에 의해 보정된 인구수이다. 5~84세에 대해서는 D_x 와 P_x 는 Beers 공식에 의해 보정된 값을 나타낸다.

$$q_x = \frac{D_x}{P_x + \frac{1}{2}D_x}$$

85~99세 사망확률의 경우에는 의료보험(medicare)자료를 이용하여 계산하고 있다. 의료보험은 등록 시 나이를 증명해야 하므로 연령 정보가 좀 더 정확하며, 85세 이상에서는 의료보험 자료를 이용한 사망확률이 인구동태통계나 센서스 자료에 비해 좀 더 정확하고 합리적이라는 판단 때문이었다. 고연령의 사망확률은 아래의 식과 같이 계산할 수 있다.

$$q_x = q_{x-1} \cdot e^{k_x},$$

$$k_x = \ln(q_x) - \ln(q_{x-1})$$



여기서 k_x 는 연령별 사망률 변화율을 의미하며 의료보험 자료로부터 구한다.

3) 생명표

생명표 작성은 아래와 같이 일련의 작성방법을 따르고 있다. 이때 정지인구 L_x 가 0이 되는 나이(보통 110~120세)까지 계산함으로써 ${}_{\infty}L_{100}$ 을 구하고 있다. 100세 이상의 고연령 사망확률은 앞서 설명한 산출방법에 의해 연장시키며, $k_x = k_{85} + (x - 85)s$, s 는 k_x 변화의 기울기, $k_{85} = [l_n(q_{88}/q_{81})]/7$ 이다.

$$(\text{생존자수}) \quad l_x = l_{x-1}(1 - q_{x-1}), \quad l_0 = 100,000$$

$$(\text{사망자수}) \quad d_x = l_x - l_{x+1} = l_x q_x, \quad {}_{\infty}q_{100} = 1.0$$

$$(\text{정지인구}) \quad L_x = \frac{1}{2}(l_x + l_{x+1}) = l_x - \frac{1}{2}d_x, \quad L_0 = fl_0 + (1-f)l_1$$

$$(\text{총생존년수}) \quad T_x = \sum_{t=0}^{\infty} L_{x+t}$$

$$(\text{기대여명}) \quad e_x = \frac{T_x}{l_x}$$

나. 10년 주기 전국 및 주(state)별 완전생명표

미국의 10년 주기 완전생명표는 센서스 해와 그 전후에 해당되는 3년을 관찰기간으로 두고 있으며, 상한 연령은 100세이다. (1900-1902)생명표 작성을 시작으로 센서스가 수행되는 10년마다 작성되며, 전국과 주 수준에서 성(남/여)과 인종(백인/백인외/흑인)별로 완전생명표를 작성하고 있다. 단, 주의 성별 인종별 사망자수가 최소 700명 이상인 경우에 대해서만 작성한다. (1989-1991)생명표 작성법에 대하여 NCHS(1998)의 내용을 정리하였다.

1) 기초자료

생명표 작성의 기초자료인 인구수는 센서스 자료를 이용하고 있다. 그러나 공식적인 센서스 자료는 아니며, 연령과 인종의 정의를 인구동태통계에 맞게 센서스 국(U.S.

Bureau of Census)에서 조정한 자료이다. 사망자의 경우에는 연령미상에 대한 보정을 먼저 수행한다. (1989-1991)생명표를 위한 3년의 관찰기간 동안 총 사망자는 650만명이었는데 그 중 연령미상이 1,700명으로 연령미상을 위해 조정되는 규모는 그리 많지 않은 편이다. 연령미상인 그룹의 연령분포는 보고된 사망자의 연령분포와 같다고 가정하고 연간 완전생명표에서의 방법과 동일하게 각 연령 그룹에 사망자수를 비례하여 배분한다.

2) 사망확률

2세 이하에 대해서는 생명표 상의 사망자수와 생존자수가 먼저 계산된다. x 세와 $x+t$ 세 사이의 사망자수(${}_t d_x$)는 아래의 수식과 같이 계산되는데, 여기서 ${}_t D_x$ 는 1989~91년 동안 x 세와 $x+t$ 세 사이에 관측된 사망자수이고, l_0 는 100,000이며 ${}_t E_x$ 는 <표 2-6>과 같다. 이는 1987, 88, 89, 90, 91의 출생자 분포가 균일하다는 가정에 기초하여 구한다.

$${}_t d_x = \frac{l_0 {}_t D_x}{{}_t E_x}$$

이와 같이 계산된 사망자수를 이용하여 2세까지의 생존자수를 $l_{x+t} = l_x - {}_t d_x$ 로 구할 수 있다. 2세 이하의 사망확률은 아래와 같이 계산한다.

$$q_0 = 1 - \frac{l_1}{l_0}, \quad q_1 = 1 - \frac{l_2}{l_1}$$

또한 전국 수준에서 생의 첫해에 대하여 0, 1, 7, 28일별 사망확률을 아래의 식에 의해 계산하지만, 주별로는 공표되지 않는다.

$${}_t q_x = 1 - \frac{l_{x+t}}{l_x}$$

〈표 2-6〉 2세 이하의 사망자수 계산을 위한 분모

연령구간		${}_tE_x$
0-1일	$\frac{1}{730}(B_{1988} + 730B_{1989} + 730B_{1990} + 729B_{1991})$
1-7일	$\frac{1}{730}(8B_{1988} + 730B_{1989} + 730B_{1990} + 722B_{1991})$
7-28일	$\frac{1}{730}(35B_{1988} + 730B_{1989} + 730B_{1990} + 695B_{1991})$
28-365일	$\frac{1}{730}(393B_{1988} + 730B_{1989} + 730B_{1990} + 337B_{1991})$
1-2년	$\frac{1}{2}(B_{1988} + 2B_{1989} + 2B_{1990} + B_{1991})$

B_z 는 z 세의 출생자 수

다음으로 2~4세의 사망확률을 구하기 위해 먼저 사망률을 아래와 같이 산출한다.

$$m_x = \frac{D_x}{P_{x-1} + P_x + P_{x+1}}$$

여기서 P_x 는 작성기간 중앙 해의 x 세 인구를 나타낸다. 또한 사망확률과 사망률간의

$q_x = \frac{2m_x}{2 + m_x} = \frac{m_x}{1 + 0.5m_x}$ 와 같은 관계식에 의해 사망확률은 아래와 같이 구하였다.

$$q_x = \frac{D_x}{P_{x-1} + P_x + P_{x+1} + \frac{1}{2}D_x}$$

5~94세의 사망확률은 아래의 식에 의해 계산하였다. 이때의 P_x 와 D_x 는 5세 간격의 자료를 Beers 공식에 의해 내삽한 값이다. 내삽방법은 연간 완전생명표에서의 방법과 동일하다.



$$q_x = \frac{D_x}{3P_x + \frac{1}{2}D_x} \quad (1)$$

다음으로 85~94세에 대한 사망확률은 위의 식(1)에 의해 계산된 사망확률(q_x^C)과 의료보험자료로부터 구한 사망확률(q_x^M)을 혼합하여 아래와 같이 계산하였다.

$$q_x = \frac{1}{11}[(95-x)q_x^C + (x-84)q_x^M]$$

그리고 95세 이상의 사망확률은 의료보험 자료로부터 여러 단계를 거쳐 구해지는데, 최종적인 사망확률은 아래와 같이 근사적으로 구할 수 있다. 고연령에서 의료보험 자료가 전통적인 사망확률보다 정확하다고 알려져 있기 때문에 의료보험 자료를 활용하고 있다. 이에 대한 자세한 계산과정은 NCHS(1998)을 참고하기 바란다.

$$q_x \cong -\frac{1}{16}q_{x-1.5} + \frac{9}{16}q_{x-0.5} + \frac{9}{16}q_{x+0.5} - \frac{1}{16}q_{x+1.5}$$

3) 생명표

10년주기 생명표도 마찬가지로 일련의 생명표 작성과정을 따르고 있다. 또한 생명표 계산과정 상의 반올림에 대한 규칙을 마련해놓고 있다. l_x 와 T_x 는 정수로 q_x 는 소수점 5째 자리, 0e_x 는 소수점 2째 자리로 반올림한다. 이와 같이 반올림된 l_x 와 T_x 에 의해 마지막으로 d_x 와 L_x 가 구해진다.

$$(\text{생존자수}) \quad l_{x+1} = l_x - l_x q_x$$

$$(\text{정지인구}) \quad L_x = 0.5(l_x + l_{x+1}) : 130\sim 110\text{세}$$

$$(\text{총생존년수}) \quad T_x = T_{x+1} + L_x : 130\sim 110\text{세} \quad (T_{131} = 0)$$

$$T_{110} = l_{110} {}^0e_{110}$$



$$T_x = T_{x+1} + 0.5(l_x + l_{x+1}) : 109\sim 0\text{세}$$

$$T_x = T_{x+t} + \frac{t}{2}(l_x + l_{x+t}) : \text{생의 첫째}$$

$$* \text{0day}(t = \frac{1}{365}), \text{1day}(t = \frac{6}{365}), \text{7days}(t = \frac{21}{365}), \text{28days}(t = \frac{337}{365})$$

$$(\text{기대여명}) {}^0e_x = \frac{T_x}{l_x} : 0\sim 109\text{세}$$

$$(\text{사망자수}) d_x = l_x - l_{x+1}$$

$$(\text{정지인구}) L_x = T_x - T_{x+1}$$

또한 85세 이하의 연령에 대한 사망확률 분산 추정값과 0~109세의 기대여명에 대한 분산 추정값은 각각 다음과 같다.

$$S^2(q_x) = \frac{q_x^2(1-q_x)}{D_x^*}$$

$$S^2({}^0e_x) = \frac{l_{110}^2 S^2({}^0e_{110}) + \sum_{y=x}^{109} l_y^2 ({}^0e_{y+1} + 0.5)^2 S^2(q_y)}{l_x^2}$$

다. 사망원인 생명표

미국의 사망원인 생명표는 사망원인 제거시 생명표(Cause-elimination life tables)로서 만약 특정 사인이 제거되었을 때 코호트의 사망경험이 어떻게 될 것인지에 대해 설명하는 생명표이다. 사망원인 생명표는 (1959-61)생명표를 시작으로 3년간의 자료를 이용하여 10년마다 작성되고 있다. 생명표의 상한 연령은 100세이며, 성(남/여)과 인종(백인/흑인)별로 51개의 사망원인에 대해 5세 간격의 간이생명표가 작성되고 있다. 사망원인 생명표와 더불어 궁극적으로 사망할 확률, 특정사인 사망자수와 특정사인 제거시 증가기대여명에 대한 테이블도 공표하고 있다. 여기서는 (1989-91)사망원인 생명표에 대해 NCHS(1999)의 내용을 기초로 하여 살펴보기로 한다.

1) 기초자료

사망원인 생명표는 모든 사인에 대한 10년 주기 전국 완전생명표의 간이버전으로부터 구하게 된다. 완전생명표를 간이생명표로 전환하기 위해 ${}_n d_x = l_{x+n} - l_x$, ${}_n q_x = {}_n d_x / l_x$, ${}_n L_x = T_x - T_{x+n}$ 와 같은 식을 이용하며, ${}_∞ q_{110} = 1.0$, $l_{110} = l_{109} - d_{109}$, ${}_∞ L_{110} = T_{110} = l_{110} \cdot e_{110}$ 이다.

특정사인 i 에 의한 사망자수의 추정은 위에서 구한 n 세와 $x+n$ 세 사이의 사망자수인 ${}_n d_x$ 와 i th 사인에 대한 사망자 비율인 ${}_n r_x^i$ 의 곱으로 아래와 같이 계산한다.

$${}_n d_x^i = {}_n r_x^i \cdot {}_n d_x$$

$${}_n r_x^i = \frac{{}_n D_x^i}{{}_n D_x} \cdot \frac{D^i / D_a^i}{D / D_a}$$

여기서 ${}_n D_x^i$ 는 1989-91년 동안 관측된 특정사인 i 로 인하여 사망한 n 세와 $x+n$ 세 사이의 사망자수이고, ${}_n D_x$ 는 1989-91년 동안 관측된 n 세와 $x+n$ 세 사이의 총 사망자수이다. D_a^i 는 특정사인 i 로 사망한 나이를 아는(보고된 연령정보가 있는) 사망자수, D^i 는 특정사인 i 로 사망한 총 사망자수, D_a 는 나이를 아는 사망자수, D 는 총 사망자수를 나타낸다.

2) 사망확률

l_x 는 생명표 상의 x 세 생존자수이고, $l_x^i = \sum_{y=x}^{110} {}_n d_y^i$ 는 특정사인 i 로 사망한 x 세 이상의 사망자수의 합이다. 따라서 특정사인으로 x 세에 궁극적으로 사망할 확률 ψ_x^i 는 아래와 같이 구할 수 있다.

$$\psi_x^i = {}_∞ q_x^i = \frac{l_x^i}{l_x}$$

다음으로 특정사인 제거시 사망확률 ${}_n q_x^{(-i)}$ 은 사망확률은 특정사인 i 가 제거되었을 때 x 세까지 생존한 사람이 n 년 이내에 사망할 확률을 의미하고 아래 식과 같다.



$${}_nq_x^{(-i)} = 1 - {}_np_x^{(-i)}$$

여기서 특정사인 제거시 생존확률 ${}_np_x^{(-i)} = {}_np_x^{(1-{}_nr_x^i)}$ 이고, ${}_np_x = \frac{l_{x+n}}{l_x}$ 이다. 이와 같이 특정사인 제거시 사망확률 ${}_nq_x^{(-i)}$ 은 x 세 사람이 특정사인 i 를 제외하고 어떠한 원인에 의해서든 사망할 확률 ${}_nq_x^{-i}$ 과는 다른 개념이므로 주의해야 한다. ${}_nq_x^{-i}$ 를 수식으로 표현하면 아래와 같다.

$${}_nq_x^{-i} = \frac{l_x - l_{x+n} - {}_nd_x^i}{l_x}$$

3) 특정사인 제거시 생명표

특정사인 제거시 생명표는 다음과 같은 각각의 산출단계에 의해 작성한다.

$$(\text{특정사인 제거시 생존자수}) \quad l_{x+n}^{(-i)} = {}_np_x^{(-i)} \cdot l_x^{(-i)}, \quad l_0^{(-i)} = 100,000$$

$$(\text{특정사인 제거시 정지인구}) \quad {}_nL_x^{(-i)} = (n - {}_nf_x) \cdot l_x^{(-i)} + {}_nf_x \cdot l_{x+n}^{(-i)},$$

$${}_nf_x = \frac{n l_x - {}_nL_x}{l_x - l_{x+n}}, \quad {}_\infty L_{110}^{(-i)} = T_{110}^{(-i)} = \frac{e_{110} l_{110}^{(-i)}}{1 - {}_\infty r_{100}^i}$$

$$(\text{특정사인 제거시 총생존년수}) \quad T_x^{(-i)} = T_{x+n}^{(-i)} + {}_nL_x^{(-i)}$$

$$(\text{특정사인 제거시 기대여명}) \quad e_x^{(-i)} = \frac{T_x^{(-i)}}{l_x^{(-i)}}$$

$$(\text{특정사인 제거시 증가 기대여명}) \quad g_x^{(-i)} = e_x^{(-i)} - e_x$$

2. 캐나다 생명표

캐나다는 센서스 해 전후 3년 동안의 자료를 이용하여 완전생명표와 간이생명표를 작성하고 있다. 전국과 주(province)수준에서 성별 완전생명표를 작성하고 있지만, 인구수와 사망자수가 적은 Prince Edward Island 주에 대해서는 5세 간격의 간이생명표 형태로 작성한다. 또한 여러 개의 준주(territories)를 결합한 지역에 대해서도 간이생명표를 작성하고 있다. 전국 완전생명표의 상한 연령은 109세이고, 주는 주별 기초자료 품질에 따라 105~109세 사이에서 결정되며, Prince Edward Island 주와 준주 지역은 상한 연령이 100세 이상이다. 여기서는 (2000-2002) 완전생명표 작성법의 사례에 대해 Statistics Canada(2006)을 참고하여 정리하였다.

가. 완전생명표

1) 기초자료

생명표 작성에 필요한 사망자와 출생자에 대한 인구동태통계는 출생과 사망 등록자료로부터 추출된다. 2000~2002년까지 관측된 연령, 출생년도, 사망년도별 0~4세의 사망자수와 동일 관찰기간의 105세 이상까지의 각 세 사망자수를 기초자료로 이용하였다. 또한 인구수는 센서스 자료에 기초한 캐나다 통계청의 7월 1일 인구 추정값을 사용한다.

2) 사망확률

x 세의 사람이 $x+1$ 세가 되기 전에 사망할 확률 q_x 는 연령대에 따라서 각기 다른 방식에 의해 구하게 된다. 먼저 전국 생명표의 0~4세 사망확률은 생존확률 p_x 에 의해 다음과 같이 계산한다.

$$q_x = 1 - p_x = 1 - ({}_ap_x)({}_dp_x) = 1 - \left(\frac{P'_x}{E_x} \right) \left(\frac{E_{x+1}}{P''_x} \right)$$

여기서 P'_x 는 관찰기간 3년 동안 x 세인 사람의 수로서 2000~2002년에 대해 x 세가 되는 해의 마지막에 생존한 사람의 수라고 할 수 있다. E_x 는 2000~2002년 동안 x 세가 된 사람의 수이다. 즉, ${}_ap_x = P'_x/E_x$ 는 x 세가 된 사람이 x 세가 된 해의 마지막에 생존할 확률이라 할 수 있다. 그리고 P''_x 는 x 세가 된 해의 마지막에 생존해 있으면서 2000~2002



년 사이에 $x+1$ 번째 생일을 맞이한 사람의 수이다. 따라서 $\delta p_x = E_{x+1}/P_x''$ 는 x 세가 된 해의 마지막에 살아 있는 사람이 $x+1$ 세에 생존할 확률이라 할 수 있다. P 와 E 에 대한 자세한 산출방법은 Statistics Canada(2006)을 참고하기 바란다.

주별 생명표의 0~4세 사망확률도 전국과 동일한 방법으로 계산하여 주별 생존확률 old p_k 를 먼저 산출한다. 그런 다음 주별 생명표에서도 연령이 높아지면서 단조감소할 수 있도록 보정작업을 수행한다. 아래의 식과 같이 전국 생존확률 national p_k 를 이용하여 주별 사망확률의 기하평균 방식으로 보정된 new p_k 를 구할 수 있다. 일반적으로 $q_0 > q_1$ 이므로 1세에서 4세에 대해서만 보정하면 되고, 이와 같은 보정을 하더라도 5세의 생존자수는 보정 전과 동일하다. (2000-2002) New Brunswick 주의 완전생명표를 작성하기 위해 주별 사망확률을 실제로 보정한 결과가 <표 2-7>과 같다. 5세 생존자수는 보정 전이나 후나 동일하지만, 보정을 통하여 연령이 증가함에 따라 사망자수가 작아짐을 확인할 수 있다.

$$(\text{new } p_k) = \frac{(\text{national } p_x) \left[\prod_{k=1}^4 (\text{old } p_k) \right]^{\frac{1}{4}}}{\left[\prod_{k=1}^4 (\text{national } p_k) \right]^{\frac{1}{4}}}$$

<표 2-7> 1~4세 여자 사망확률 보정결과(New Brunswick 주)

나이(x)	old q_x 를 이용한 기대 사망자수	new q_x 를 이용한 기대 사망자수
0	368	368
1	27	34
2	9	19
3	17	14
4	25	11
5세 생존자수	99,554	99,554

다음으로 5세 이상의 사망확률에 대해서는 5세 간격의 중심나이(pivotal age)를 설정

하여 사망확률을 계산한 후 중심나이를 이용하여 그 사이 연령들에 대한 사망확률을 내삽하는 방식으로 구하고 있다. 7, 12, 17, ..., 87세의 중심나이의 사망확률은 일반적으로 알려져 있는 방법인 아래와 같은 수식에 의해 계산한다.

$$q_x = \frac{2m_x}{(2 + m_x)}$$

$$m_x = \frac{D_x/3}{P_x}$$

$$D_x = 0.216D'_x - 0.008(D'_{x-5} + D'_{x+5})$$

$$P_x = 0.216P'_x - 0.008(P'_{x-5} + P'_{x+5})$$

여기서 인구수 D_x 와 사망자수 P_x 는 King's 공식에 의해 위와 같이 산출되는데, D'_x 은 2000~2002년에 관측된 $x-2$, $x-1$, x , $x+1$, $x+2$ 세 사망자수의 합, P'_x 은 $x-2$, $x-1$, x , $x+1$, $x+2$ 세 인구 추정값의 합이다. 또한 사망확률의 분산 추정식은 아래와 같다.

$$Var(q_x) = \frac{q_x^2(1-q_x)}{(D_x^*)}$$

$$D_x^* = \frac{1}{3} \sum_{y=2000}^{2002} y \text{해 } x \text{세 사망자수}$$

또한 92, 97, 102, 107, 112세 중심나이의 사망확률은 Nagnur's 공식을 이용하여 아래와 같이 외삽되고, 분산 추정식은 다음과 같다.

$$q_x = \min\{(4q_{x-5} - 6q_{x-10} + 4q_{x-15} - q_{x-20}), 1\}$$

$$Var(q_x) = 16 Var(q_{x-5}) + 36 Var(q_{x-10}) + 16 Var(q_{x-15}) + Var(q_{x-20})$$



이와 같이 중심나이의 사망확률이 먼저 산출되면 중심나이 사이의 중간나이에 대한 사망확률을 중심나이로부터 계산한다. 13~106세의 중간나이 사망확률은 Karup-King's third-difference tangential formula에 의해 아래와 같고, 분산 추정식도 다음과 같다.

$$\begin{aligned} q_{x+1} &= \min\{(-0.064q_{x-5} + 0.912q_x + 0.168q_{x+5} - 0.016q_{x+10}), 1\}, \\ q_{x+2} &= \min\{(-0.072q_{x-5} + 0.696q_x + 0.424q_{x+5} - 0.048q_{x+10}), 1\}, \\ q_{x+3} &= \min\{(-0.048q_{x-5} + 0.424q_x + 0.696q_{x+5} - 0.072q_{x+10}), 1\}, \\ q_{x+4} &= \min\{(-0.016q_{x-5} + 0.168q_x + 0.912q_{x+5} - 0.064q_{x+10}), 1\}, \\ \text{여기서 } x &= 12, 17, 22, \dots, 102 \end{aligned}$$

$$\begin{aligned} Var(q_{x+1}) &= (0.064)^2 Var(q_{x-5}) + (0.912)^2 Var(q_x) \\ &\quad + (0.168)^2 Var(q_{x+5}) + (0.016)^2 Var(q_{x+10}) \end{aligned}$$

그리고 8~11세의 중간나이의 사망확률은 Jenkin's fifth difference osculatory non-reproducing formula로부터 아래처럼 계산된다. 이 공식은 13~106세의 중간나이에 대해 적용했던 Karup-King's formula보다 더 부드러운 커브를 만들어낸다. 5세와 6세에 대한 사망확률도 아래의 식에 의해 계산된다.

$$\begin{aligned} q_8 &= -(217q_2 - 3,892q_7 - 966q_{12} + 140q_{17} + q_{22})/4,500 \\ q_9 &= -(296q_2 - 3,056q_7 - 1968q_{12} + 220q_{17} + 8q_{22})/4,500 \\ q_{10} &= -(279q_2 - 2,124q_7 - 2,862q_{12} + 180q_{17} + 27q_{22})/4,500 \\ q_{11} &= -(208q_2 - 1,228q_7 - 3,054q_{12} + 40q_{17} + 64q_{22})/4,500 \end{aligned}$$

$$q_5 = -0.3q_3 + q_4 + 0.5q_7 - 0.2q_8$$

$$q_6 = -0.2q_3 + 0.5q_4 + q_7 - 0.3q_8$$

마지막으로 고연령에 대해서는 모형기반 추정값으로 교체해주는데 아래와 같은 Coale-Kisker 방법에 의해 사망확률을 외삽한다. 이때 M_{110} 을 남자는 1, 여자는 0.8로 가

정하고 있다. 전국은 93세부터, 주는 88세부터 앞에서 계산된 q_x 대신에 $Q_{CK}(x)$ 로 사용한다.

$$Q_{CK}(x) = 2M_{CK}(x)/(2 + M_{CK}(x)), \quad x = 87, \dots, 117$$

$$M_{CK}(x) = M_{avg} \text{EXP} \{ (x - 84) [K_{85} + (x - 85)(S/2)] \}$$

$$K_{85} = \frac{1}{4} \log(M_{86}/M_{82}), \quad S = -\frac{1}{325} [\log(M_{avg}/M_{110}) + 26K_{85}]$$

$$M_{avg} = \frac{1}{5}(M_{82} + M_{83} + M_{84} + M_{85} + M_{86})$$

$$M_x = D_x^*/P_x, \quad x = 82, 83, \dots, 86$$

나. 간이생명표

1) 기초자료

Prince Edward Island 주와 북쪽 준주지역은 인구수가 적기 때문에 22개 연령그룹(0, 1-4, 5~9, ..., 95-99, 100+)의 간이생명표를 작성하고 있다. (2000-2002)간이생명표 작성을 위해서는 1999~2002년의 출생자 수, 2000~2002년 사망자수, 2001 인구 센서스에 기초한 2001년 7월 1일 인구 추정값이 기초자료로 필요하다.

2) 사망확률

사망확률을 구하기 위해 먼저 2001년 인구수 D_x 와 2000~2002년의 사망자수 P_x 로부터 사망률 $m_x = D_x/3P_x$ 를 산출한다. 이와 같은 사망률을 이용하여 연령그룹별로 사망확률을 계산한다.

0세 사망확률은 0세의 분리계수 F_0 , 1999~2001년 출생자수 B_{99-01} 와 2000~2002년 출생자수 B_{00-02} 로부터 구한다.

$${}^1q_0 = \frac{D_0}{[B_{99-01}(F_0) + B_{00-02}(1 - F_0)]}$$



다음으로 1~4세 사망확률은 Greville의 방법으로 아래와 같이 계산한다. 여기서 w_1 은 1~4세 사이의 나이간격을 의미하므로 4의 값을 갖는다.

$${}_4q_1 = \frac{m_1}{\beta_1},$$

$$\beta_1 = \left(\frac{1}{w_1} \right) + m_1 \left[(1 - F_1) + \left(\frac{w_1}{12} \right) (m_1 - k) \right], \quad k = \left(\frac{1}{45} \right) \ln \left(\frac{m_{85}}{m_{40}} \right)$$

5-9세, 10-14세, ..., 95-99세 사망확률도 아래와 같이 구하며 분산추정식도 다음과 같다. 또한 w_x 는 5이다.

$${}_5q_x = \frac{m_x}{\beta_x},$$

$$\beta_x = \left(\frac{1}{w_x} \right) + m_x \left[0.5 + \left(\frac{w_x}{12} \right) (m_x - k) \right], \quad k = \left(\frac{1}{45} \right) \ln \left(\frac{m_{85}}{m_{40}} \right)$$

$$Var(q_x) = \frac{{}_nq_x^2 (1 - {}_nq_x)}{D_x}$$

마지막으로 100세 사망확률 $q_{100} = 1$ 이고, 생명표 작성은 일련의 방법과 동일하다.

제4절 개선방안 및 실증분석

1. 시도 생명표

가. 관찰기간과 보정방법

생명표는 설정된 관찰기간에 맞추어 인구수와 사망자수에 대한 기초자료를 수집하고 기초 사망률과 보정된 사망확률을 산출한 후 작성하게 되므로, 생명표 작성 전 단계에서 고려할 사항은 관찰기간과 보정된 사망확률 산출 방법에 있다고 할 수 있다. 일반생명표는 해당연도를 관찰기간으로 하여 작성하는 것이 일반적이지만, 시도 생명표와 같이 보다 세분화된 생명표 작성을 위해서는 어느 정도 충분한 자료가 필요하며, 3~5년 정도를

고려할 수 있다. 사망확률 보정방법은 여러 가지가 있으며, 기초자료로부터 산출된 사망 확률에 이동가중평균을 적용하여 보정하는 Greville 방법과 기초자료인 인구수와 사망자 수를 보정한 후 사망확률을 산출하는 Beers 방법 등이 있다. 3절에서 살펴본 바와 같이 미국이나 캐나다의 경우 지역 생명표는 어느 정도의 사망자수를 만족하는 지역에 한하여 3년간 자료를 이용하고 있고, 전국수준과 비교하기 위해서 동일한 기간에 대한 전국 단위 생명표를 작성하고 있다. 사망확률 보정방법은 우리나라의 경우 Greville 9차항 방법을 적용하고 있으며, 미국은 Beers 방법, 캐나다는 Karup-King's 방법, Jenkin's 방법 등을 적용하고 있다.

우리나라 시도 생명표는 전국 단위 일반생명표와 동일한 방법에 의해 각 세별 완전 생명표로 작성된 후 5세 구간으로 재작성되어 간이생명표로 제공되고 있다. 시도 생명표의 기초자료인 16개 시도의 단년치 자료는 지역에 따라 차이가 있지만 생명표를 작성하기에 연령별 사망자수가 부족해서 각 세 기초자료를 이용하여 만들어진 생명표 사망 확률은 매우 불안정한 패턴을 보이고 있다. 따라서 시도별 사망확률에 대해 좀 더 안정성을 확보할 수 있는 방안으로 관찰기간(1년, 3년)과 Greville 보정방법과 Beers 보정방법을 중심으로 검토하고자 한다. Greville 방법과 Beers 방법은 모두 적용하기에 용이하며 보정계수의 불변성을 만족하는 장점을 가지고 있다.

Beers 방법은 5세 단위 기초자료를 묶은 후 보정된 각 세 자료로 뿌려주는 방법으로 원 자료와 보정된 자료의 총 합이 같게 되는 장점이 있으며, 적용을 위해서는 <표 2-5>의 보정계수와 아래의 식이 필요하다.

$$\begin{aligned}
 (0\sim 4\text{세}) \quad N_{0+k} &= C_{k,0\ 5}N_0 + C_{k,5\ 5}N_5 + C_{k,10\ 5}N_{10} + C_{k,15\ 5}N_{15} + C_{k,20\ 5}N_{20} \\
 (5\sim 9\text{세}) \quad N_{5+k} &= C_{k,0\ 5}N_0 + C_{k,5\ 5}N_5 + C_{k,10\ 5}N_{10} + C_{k,15\ 5}N_{15} + C_{k,20\ 5}N_{20} \\
 (10\sim 94\text{세}) \quad N_{x+k} &= C_{k,x-10\ 5}N_{x-10} + C_{k,x-5\ 5}N_{x-5} + C_{k,x\ 5}N_x + C_{k,x+5\ 5}N_{x+5} \\
 &\quad + C_{k,x+10\ 5}N_{x+10} \\
 (95\text{세 이상}) \quad N_{95+k} &= C_{k,80\ 5}N_{80} + C_{k,85\ 5}N_{85} + C_{k,90\ 5}N_{90} + C_{k,95\ 5}N_{95} \\
 &\quad + C_{k,100\ 5}N_{100}
 \end{aligned} \tag{2}$$

여기서 N_{x+k} 는 $(x+k)$ 세의 인구수 또는 사망자수이며, $C_{k,x}$ 는 Beers 보정계수를 나타낸다.

Greville 방법은 이동가중평균을 이용한 보정방법으로 사용되는 항의 수에 따라 5차,



7차, 9차, 11차, 13차 등이 있으며, 다음과 같다. Greville 방법은 차수가 높을수록 평활도는 높아지고, 적합도는 낮아지게 되므로 적절한 차수를 선택하여 조절하는 것이 가능하다는 장점이 있다.

(5차항)

$$q_x = -0.073427q_{x-2} + 0.2937063q_{x-1} + 0.5594406q_x + 0.2937063q_{x+1} - 0.073427q_{x+2}$$

(7차항)

$$q_x = -0.058741q_{x-3} + 0.0587413q_{x-2} + 0.2937063q_{x-1} + 0.4125874q_x + 0.2937063q_{x+1} \\ + 0.0587413q_{x+2} - 0.058741q_{x+3}$$

(9차항)

$$q_x = -0.040724q_{x-4} - 0.009873q_{x-3} + 0.11847q_{x-2} + 0.266557q_{x-1} + 0.331140q_x \\ + 0.266557q_{x+1} + 0.11847q_{x+2} - 0.009873q_{x+3} - 0.040724q_{x+4}$$

(11차항)

$$q_x = -0.027864q_{x-5} - 0.026792q_{x-4} + 0.035723q_{x-3} + 0.141267q_{x-2} + 0.2386932q_{x-1} \\ + 0.277945q_x + 0.238693q_{x+1} + 0.1412674q_{x+2} + 0.0357228q_{x+3} - 0.026792q_{x+4}$$

(13차항)

$$q_x = -0.01935q_{x-6} - 0.027864q_{x-5} - 0.000000q_{x-4} + 0.0654918q_{x-3} + 0.1473565q_{x-2} \\ + 0.2143367q_{x-1} + 0.2400572q_x + 0.2143367q_{x+1} + 0.1473565q_{x+2} + 0.0654918q_{x+3} \\ - 0.000000q_{x+4} - 0.027864q_{x-5} - 0.01935q_{x-6} \quad (3)$$

나. 실증분석

분석을 위해 관찰기간으로 1년과 3년을 고려하고 보정방법은 Greville 방법과 Beers 방법을 적용하여 비교하였다. 자료는 관찰기간이 1년인 경우에는 2008년 시도별 인구수와 사망자수이며, 관찰기간이 3년인 경우에는 2007~2009년 시도별 인구수와 사망자수를 사용하였다. 관찰기관과 보정방법을 고려하여 다음의 4가지 방법을 적용하고 산출된 로그사망확률과 기대수명에 대해 현재 시도 생명표와 비교하였다. Greville 보정의 차수는 안정적인 사망확률을 산출하기 위하여 평활도가 높은 13차항 보정방법을 적용하였다.

(방법1) 2008년 기초자료와 Greville(13차항) 보정

(방법2) 2008년 기초자료와 Beers 보정

(방법3) 2007~2009년 기초자료와 Greville(13차항) 보정

(방법4) 2007~2009년 기초자료와 Beers 보정

각 방법에 따라 전국단위와 16개 시도의 각 세별 사망확률을 산출하고 생명표를 작성하였다. 2007~2009년 3년 자료를 이용하는 (방법3)과 (방법4)의 경우의 사망확률은 다음과 같이 구하였다.

$$(0\text{세}) \quad d_0 = D_0/B, \quad B = \frac{1}{2}(B_{2006} + B_{2009}) + B_{2007} + B_{2008}, \quad q_0 = 1 - l_0/l_1$$

$$(1\sim 99\text{세}) \quad m_x = D_x/P_x, \quad q_x = m_x/(1 + 0.5m_x)$$

여기서 B_t 는 t -년도의 출생자수이며, P_x 와 D_x 는 각각 2007~2009년 동안 x 세 인구수 및 사망자수이다. 구해진 연령별 사망확률을 각 방법대로 보정하였으며, 85세 이상에 대해서는 최고연령 130세의 사망확률을 남자는 0.8, 여자는 0.6으로 가정하고 Coale and Kisker(1990) 방법을 적용하여 보정하였다. 85세 이상에 대한 보정방법은 다음과 같이 표현된다.

$$k_x = k_{85} + (x - 85) \cdot s, \quad k_{85} = \ln(q_{85}/q_{84})$$

$$s = - \left[\ln\left(\frac{q_{84}}{q_{130}}\right) + 45k_{85} \right] / 1035$$

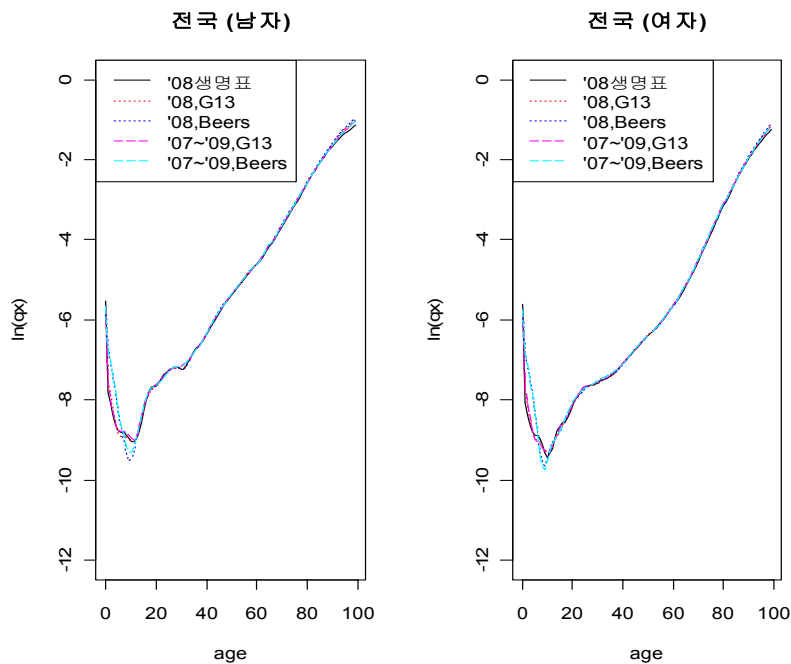
$$q_x = q_{x-1} \times e^{k_x}.$$

남녀 각각에 대해 전국 및 16개 시도별로 각 방법에 따른 로그사망확률을 [그림 2-1], [그림 2-2], [그림 2-3]에 수록하였다. [그림 2-1]의 전국 로그사망확률을 보면, 남녀 모두 각 방법에 의한 차이가 크지 않다. 다만, Beers 보정방법의 저연령층 사망확률이 Greville 방법의 비해서 최저점이 좀 더 뾰족함을 알 수 있다. 이와 같은 현상은 시도에서도 마찬가지로 나타난다. [그림 2-2]에서 서울지역 남자의 경우 2008 생명표와 단년치 Greville

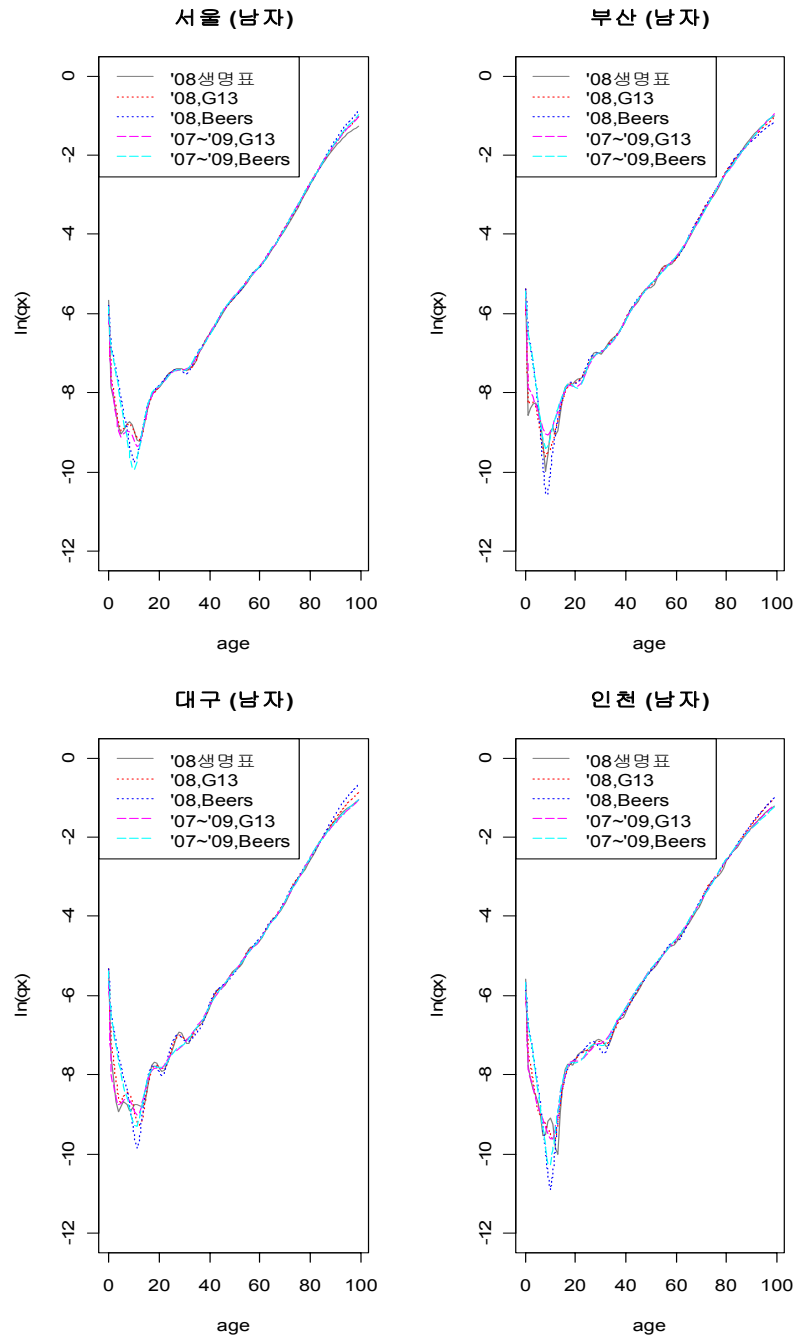


방법의 저연령층 로그사망확률이 2개의 봉우리를 만들고 있다. 3년치 Greville 방법이 그
에 비해 다소 완만해진 형태를 보이고 있으나 2개의 봉우리는 여전히 존재한다. 반면
Beers방법은 단년치와 3년치 모두 한 개의 봉우리만을 갖는다. 이는 경남의 사례에서도
더 확연히 드러난다.

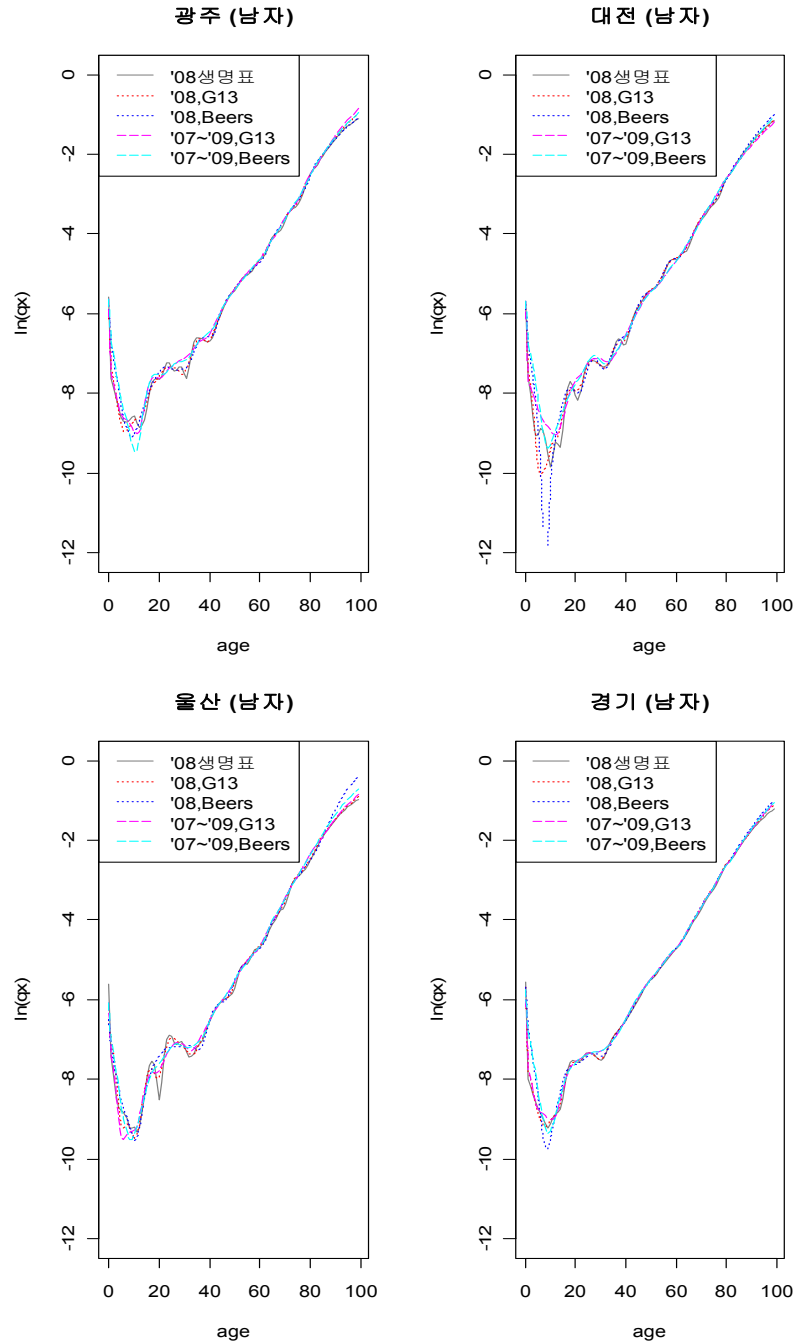
그 이외의 연령층에서는 보정방법에 의한 차이는 뚜렷하게 나타나지 않지만 관찰기
간에 의한 차이를 확인할 수 있다. 인구수와 사망자수가 적어서 생명표 상의 로그사망확
률이 가장 불안정한 제주도의 경우 단년치를 이용한 경우보다 3년의 자료를 이용하였을
때 훨씬 안정되는 것을 알 수 있다. 또한 대전의 경우 2008년 남자 8세의 사망자가 없고
주변 연령의 사망자수도 적어 Beers 보정 후 8세의 사망확률이 음의 값을 보였으나 3년
의 관찰기간을 두었을 때는 이러한 문제를 해결할 수 있다. 로그사망확률이 남자보다 더
불안정한 여자에 대해서도 [그림 2-3]에서 동일한 결과를 확인할 수 있다.



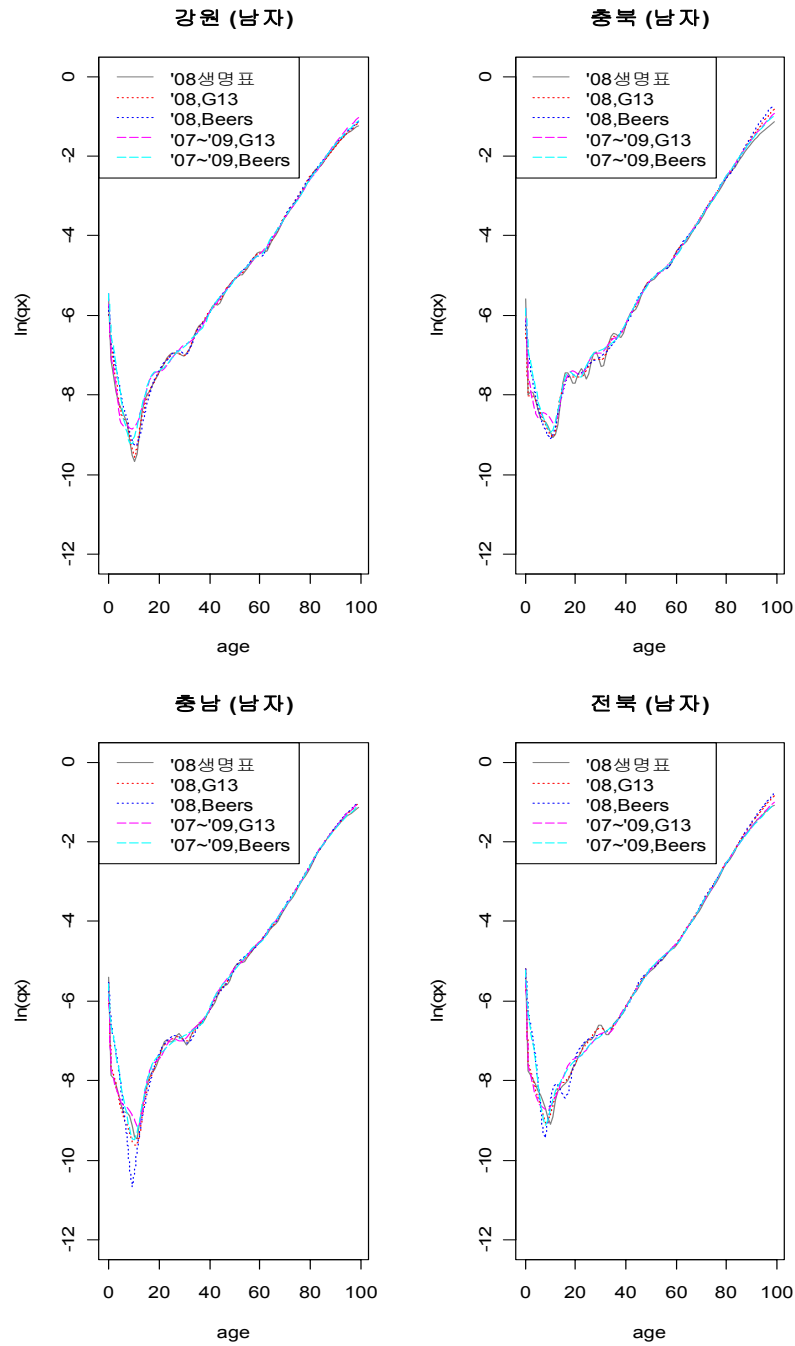
[그림 2-1] 전국 로그사망확률



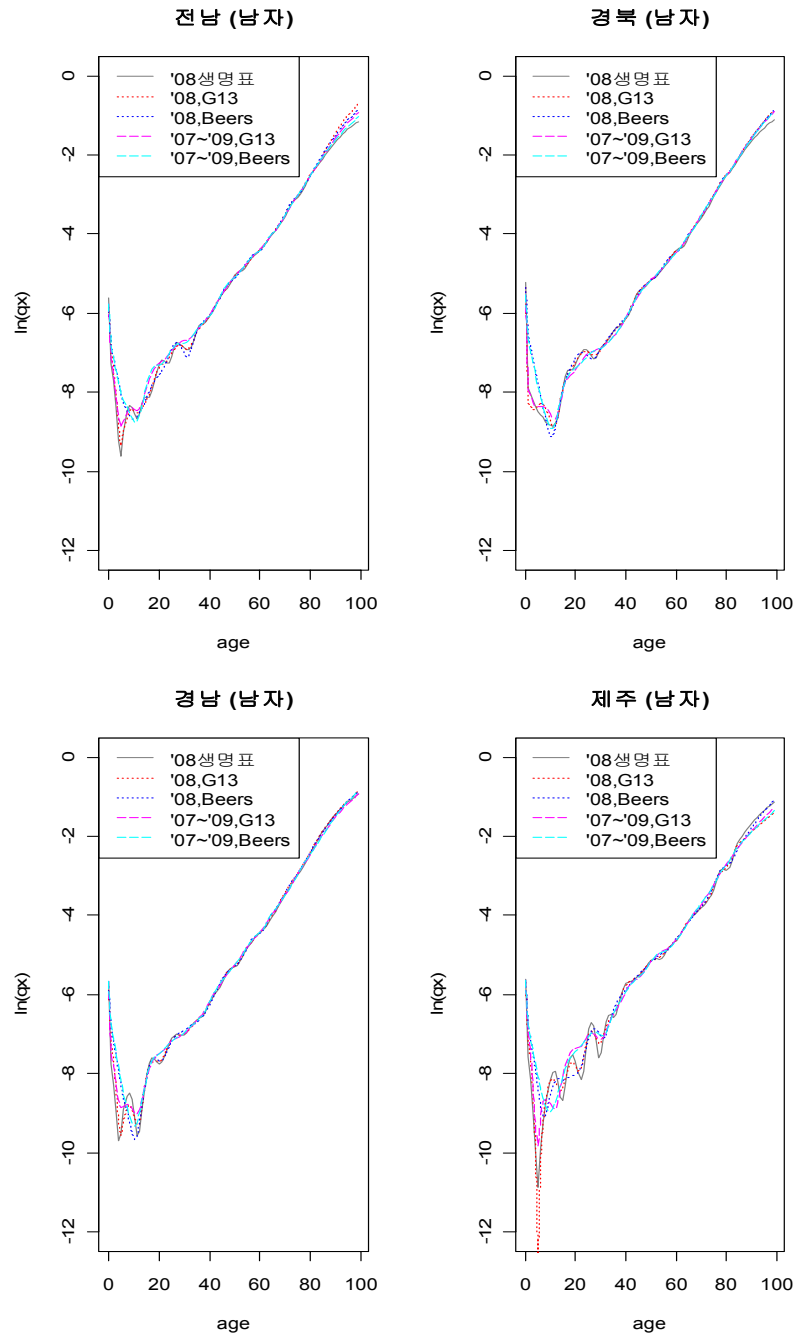
[그림 2-2] 시도별 로그사망확률 - 남자



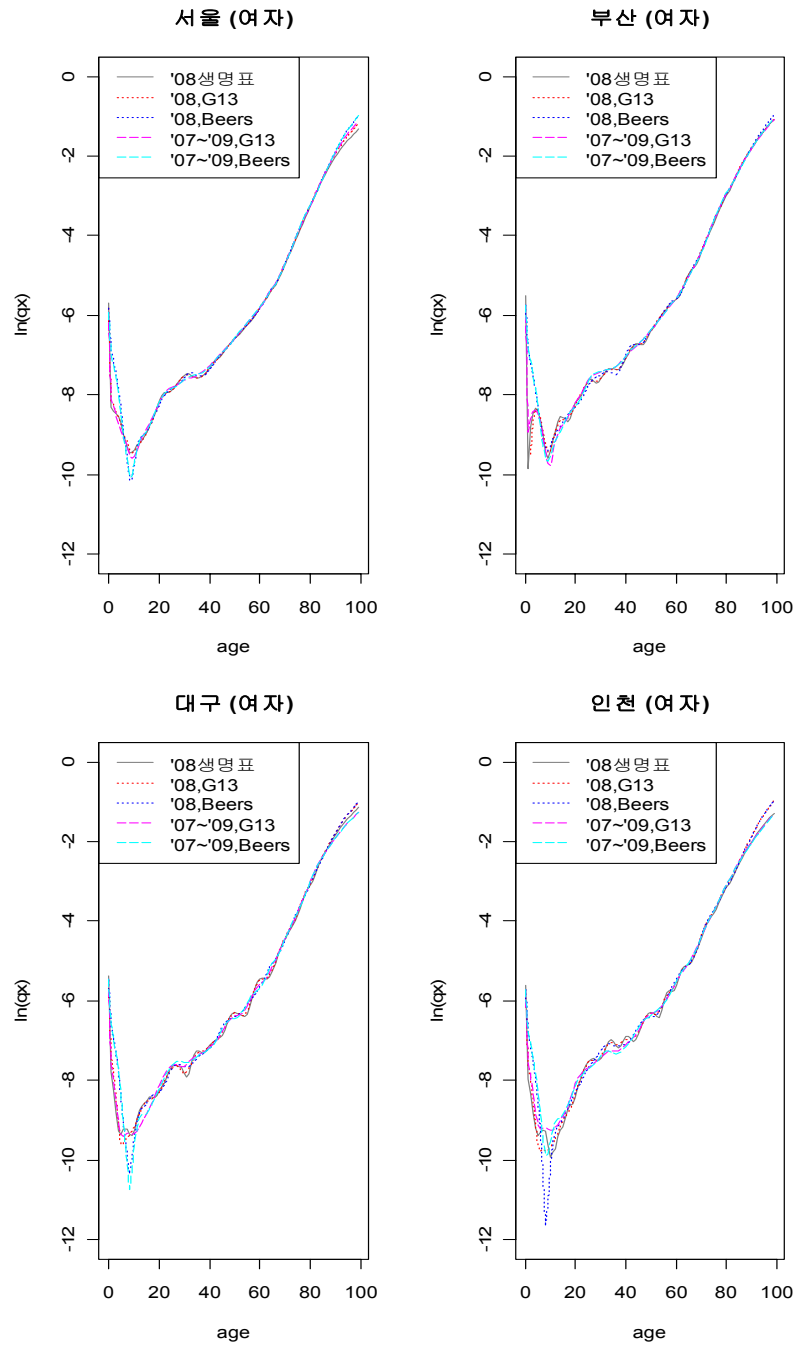
[그림 2-2] 시도별 로그사망확률 - 남자(계속)



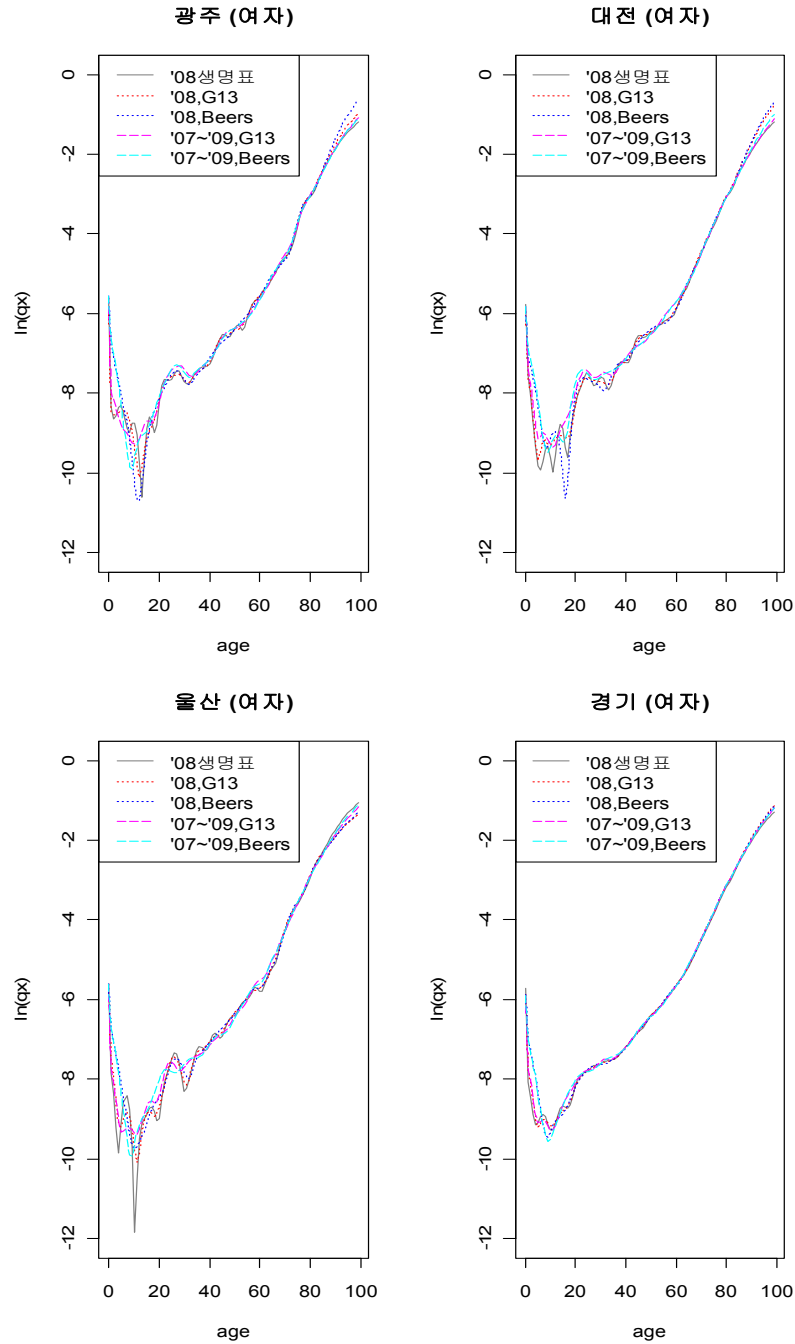
[그림 2-2] 시도별 로그사망확률 - 남자(계속)



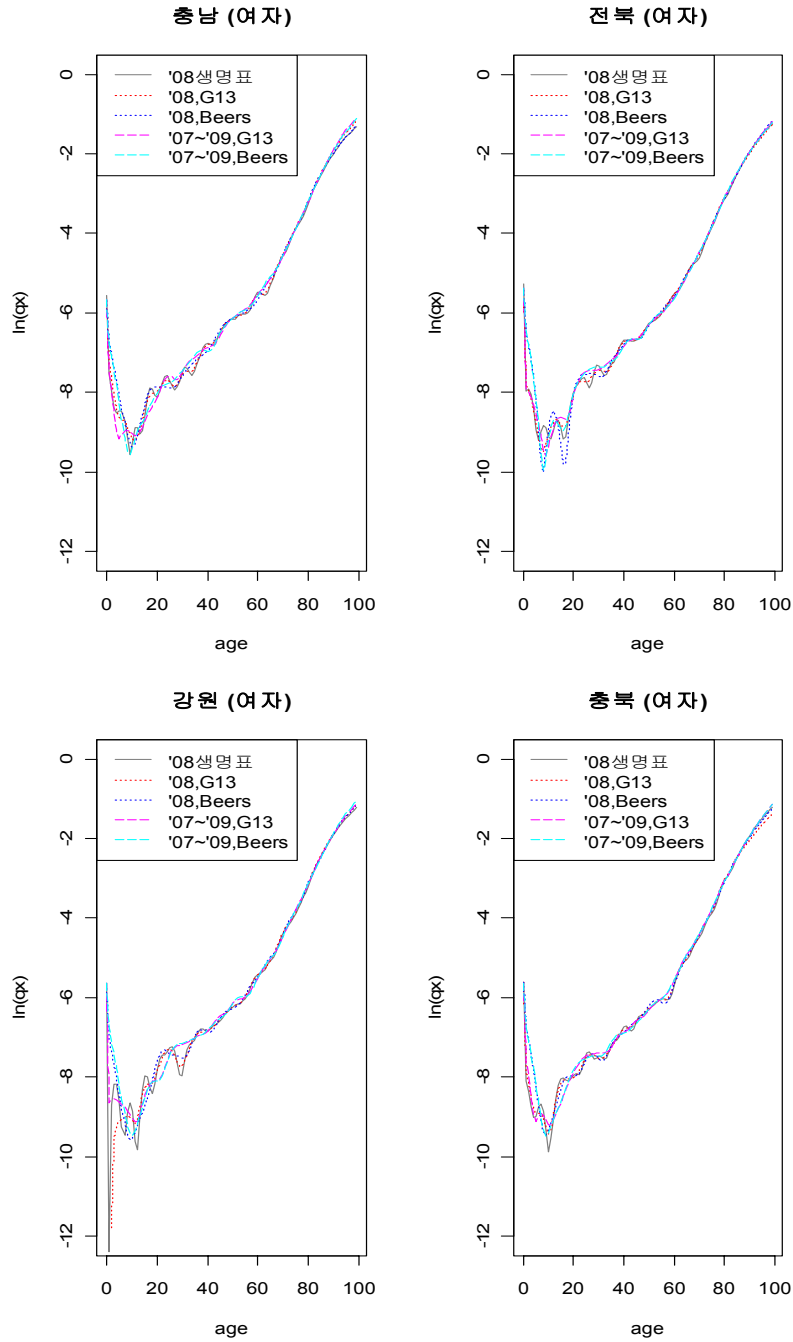
[그림 2-2] 시도별 로그사망확률 - 남자(계속)



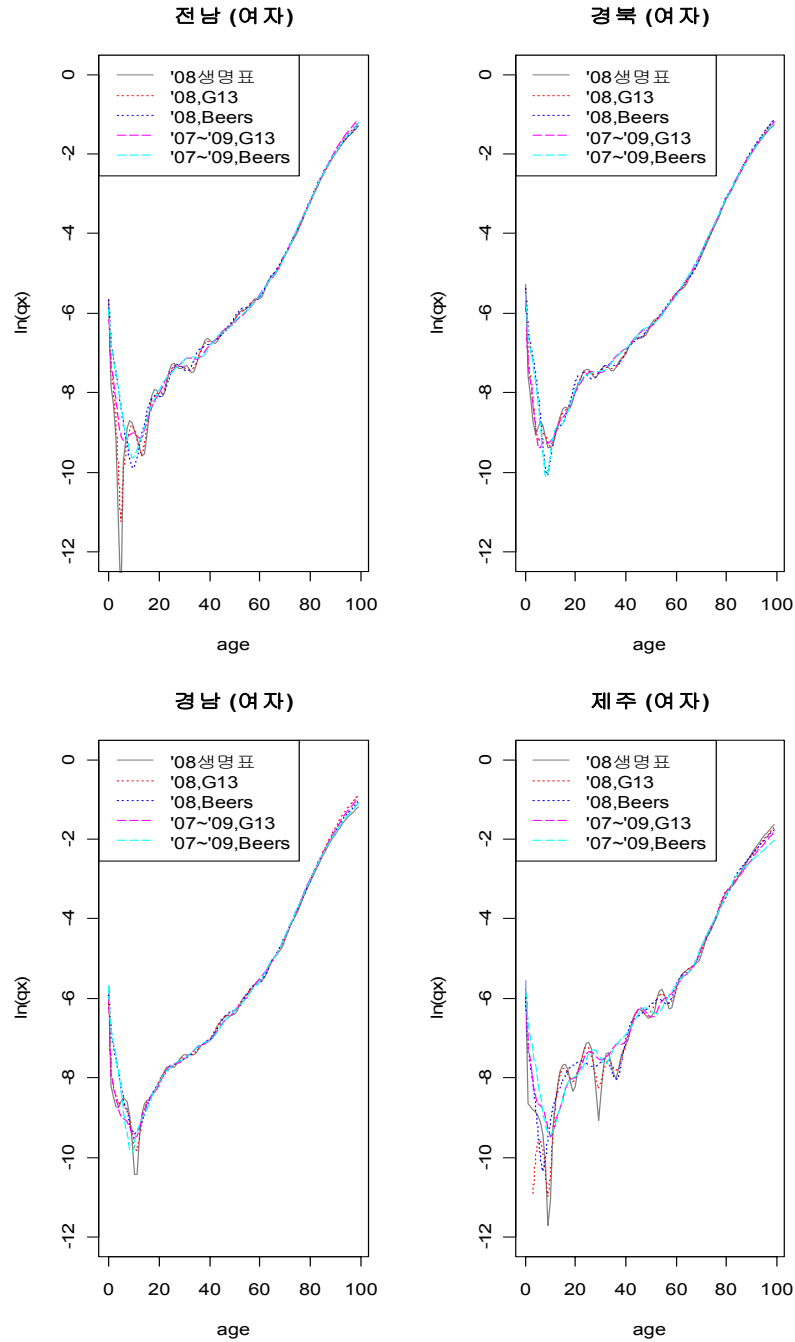
[그림 2-3] 시도별 로그사망확률 - 여자



[그림 2-3] 시도별 로그사망확률 - 여자(계속)



[그림 2-3] 시도별 로그사망확률 - 여자(계속)



[그림 2-3] 시도별 로그사망확률 - 여자(계속)

〈표 2-8〉 기대수명

지역	남자					여자				
	2008 생명표	관찰기간 '08		관찰기간 '07~'09		2008 생명표	관찰기간 '08		관찰기간 '07~'09	
		G13	Beers	G13	Beers		G13	Beers	G13	Beers
전국	76.54	76.27	76.12	76.32	76.18	83.29	83.00	82.84	83.00	82.87
서울	78.49	78.20	77.99	78.14	77.98	84.45	84.10	83.81	84.05	83.83
부산	75.23	75.07	74.90	75.28	75.08	82.06	81.95	81.77	81.96	81.82
대구	76.05	75.76	75.52	76.09	75.86	82.70	82.44	82.31	82.58	82.42
인천	76.67	76.33	76.21	76.34	76.21	83.07	82.54	82.42	82.78	82.69
광주	76.71	76.49	76.43	76.25	76.15	82.83	82.44	82.07	82.60	82.44
대전	77.13	76.95	76.78	77.00	76.82	83.19	82.65	82.53	82.90	82.67
울산	76.09	75.92	75.88	75.63	75.51	81.94	81.98	81.82	81.90	81.69
경기	77.44	77.19	76.99	77.18	77.04	83.55	83.20	83.12	83.23	83.12
강원	75.19	74.93	74.81	74.92	74.80	82.68	82.49	82.33	82.43	82.22
충북	75.41	75.13	75.05	75.03	74.95	82.72	82.73	82.44	82.36	82.20
충남	76.02	75.79	75.62	75.65	75.51	83.35	82.98	82.98	82.91	82.80
전북	75.50	75.16	74.92	75.35	75.17	82.78	82.74	82.44	82.61	82.43
전남	74.94	74.52	74.47	74.53	74.46	83.15	82.84	82.72	82.83	82.84
경북	75.02	74.73	74.54	74.92	74.77	82.89	82.58	82.36	82.57	82.42
경남	74.96	74.67	74.61	74.82	74.70	82.62	82.11	82.13	82.17	82.13
제주	76.49	76.60	76.38	76.40	76.32	85.43	85.53	85.43	85.59	85.77

각 방법에 따른 기대수명을 <표 2-8>에 수록하였다. 2008년 생명표와 비교해 보면 Greville 13차항이나 Beers 방법에 의한 기대수명이 지역과 성별에 따라 차이가 있으나 대체적으로 낮다. 관찰기간이 1년인 경우보다는 3년인 경우가, Greville 보정보다는 Beers 보정이 기대수명을 다소 낮게 산출하는 경향을 보인다.

보정방법에 대한 적합도(Fitness)와 평활도(Smoothness) 평가를 위하여 각각 MSE(mean squared error)와 3차 차분 합을 이용하는 것이 보편적인 것으로 알려져 있으며, 다음과 같다(London, 1981).



$$F = \sum_x^N (q_x^a - q_x^b)^2 / N$$

$$S = \sum_x^N |q_{x+3}^b - 3q_{x+2}^b + 3q_{x+1}^b - q_x^b|$$

여기서 q_x^a 는 보정대상이 되는 값(관측치)이며, q_x^b 는 보정된 값을 나타낸다.

각 방법에 따른 사망확률 적합도를 <표 2-9>와 [그림 2-4]에 나타내었다. 적합도 평가를 위한 기준값은 각 세별 사망자수와 인구수 자료를 이용하여 계산한 각 세별 사망확률로 아무런 보정과정을 거치지 않은 실제 관측 값이며, 방법과 관찰기간에 따라 각각 1, 3년에 대한 자료를 이용하여 계산하였다. 2008 생명표 남자의 경우 전국의 적합도 0.00003을 기준으로 각 시도별 수준을 비교해보면, 서울과 경기도는 각각 0.00007, 0.00008의 수준으로 전국에 근접한 수준인 반면 대전 0.00322, 울산 0.005, 제주 0.00222로 다른 지역에 비해 적합도가 떨어진다. 3년의 관찰기간에 대해서 남자의 경우 울산을 제외한 시도의 적합도가 전국과 유사한 수준으로 향상됨을 알 수 있다. 여자의 경우 적합도 수준이 남자에 비해 높은 편으로 각 방법별, 시도별로 적합도의 편차가 크지 않다. 또한 관찰기간과 관계없이 Greville 방법이 Beers 방법에 비해 남녀 모두 적합도가 높은 것으로 나타났다.

그리고 각 방법에 대한 사망확률 평활도를 시도별로 산출하고 <표 2-10>과 [그림 2-4]에 나타내었는데, 3차 차분합 S 값이 작을수록 평활도가 높은 것으로 해석하면 된다. 남자의 경우 2008년 생명표 울산과 제주의 평활도 수준이 다른 시도에 비해 떨어지는데, [그림 2-2]에서 보는 바와 같이 사망확률이 불안정한 형태이므로 안정화를 위한 보정이 필요한 것으로 판단된다. 남녀 모두 2008년 생명표보다는 Greville과 Beers방법으로 보정을 한 경우가, 관찰기간 1년보다는 3년이 평활도가 높아짐을 확인할 수 있다. 16개 시도 평균수준에서 볼때 3년의 관찰기간인 Greville 13차항 보정을 한 경우의 평활도가 가장 높다. 또한 관찰기간을 1년으로 했을 때는 Beers의 평활도가 높으나, 관찰기간을 3년으로 하게 되면 Greville 방법의 평활도가 높음을 알 수 있다.

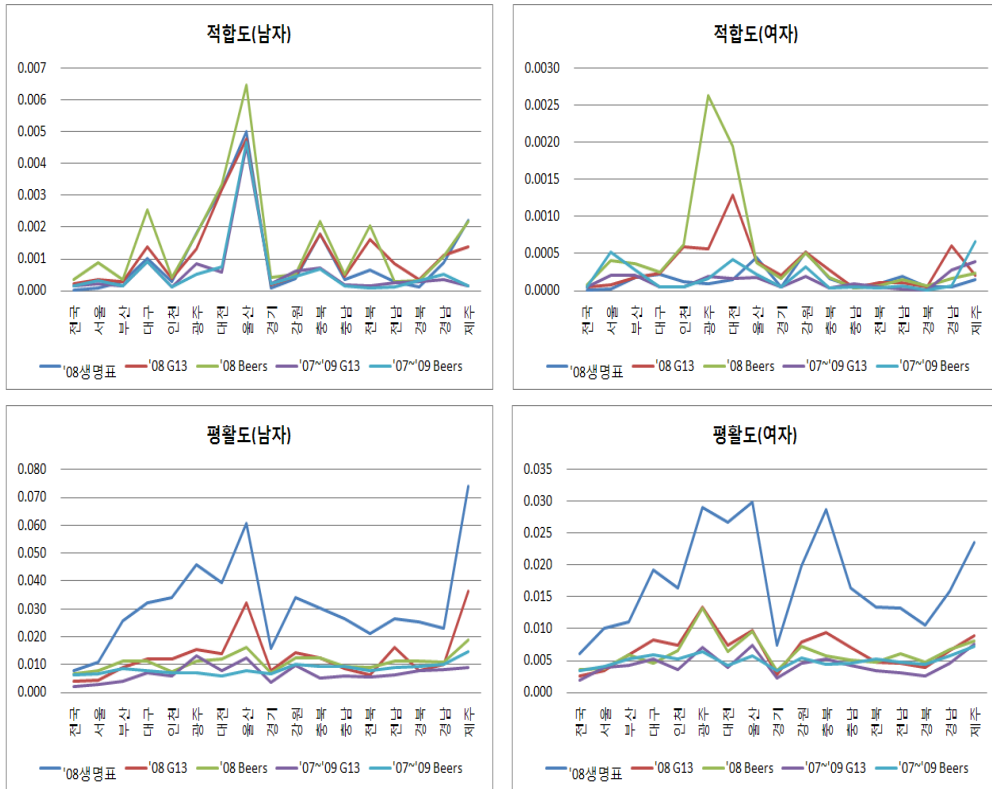
〈표 2-9〉 사망확률 적합도

지역	남자					여자				
	2008 생명표	관찰기간 '08		관찰기간 '07~'09		2008 생명표	관찰기간 '08		관찰기간 '07~'09	
		G13	Beers	G13	Beers		G13	Beers	G13	Beers
전국	0.00003	0.00021	0.00035	0.00015	0.00015	0.00001	0.00005	0.00008	0.00005	0.00005
서울	0.00007	0.00035	0.00088	0.00021	0.00032	0.00003	0.00008	0.00040	0.00021	0.00053
부산	0.00027	0.00027	0.00036	0.00016	0.00016	0.00017	0.00018	0.00036	0.00021	0.00028
대구	0.00102	0.00139	0.00255	0.00091	0.00093	0.00022	0.00024	0.00025	0.00004	0.00005
인천	0.00028	0.00040	0.00042	0.00011	0.00011	0.00012	0.00059	0.00063	0.00005	0.00004
광주	0.00183	0.00132	0.00179	0.00086	0.00053	0.00009	0.00056	0.00263	0.00020	0.00017
대전	0.00322	0.00317	0.00334	0.00057	0.00076	0.00015	0.00129	0.00194	0.00017	0.00043
울산	0.00500	0.00478	0.00649	0.00455	0.00467	0.00046	0.00040	0.00037	0.00017	0.00023
경기	0.00008	0.00018	0.00042	0.00021	0.00023	0.00005	0.00020	0.00016	0.00005	0.00006
강원	0.00039	0.00046	0.00053	0.00063	0.00044	0.00052	0.00052	0.00051	0.00020	0.00033
충북	0.00179	0.00178	0.00220	0.00073	0.00070	0.00016	0.00028	0.00017	0.00004	0.00004
충남	0.00036	0.00045	0.00053	0.00019	0.00014	0.00005	0.00003	0.00004	0.00009	0.00005
전북	0.00065	0.00160	0.00204	0.00014	0.00010	0.00009	0.00011	0.00006	0.00005	0.00003
전남	0.00028	0.00084	0.00030	0.00025	0.00011	0.00019	0.00011	0.00014	0.00003	0.00006
경북	0.00013	0.00034	0.00033	0.00028	0.00030	0.00006	0.00005	0.00007	0.00001	0.00001
경남	0.00088	0.00112	0.00109	0.00035	0.00052	0.00005	0.00060	0.00017	0.00029	0.00007
제주	0.00222	0.00140	0.00219	0.00014	0.00014	0.00015	0.00021	0.00024	0.00040	0.00066
시도 평균	0.00115	0.00124	0.00159	0.00064	0.00063	0.00016	0.00034	0.00051	0.00014	0.00019

〈표 2-10〉 사망확률 평활도

지역	남자					여자				
	2008 생명표	관찰기간 '08		관찰기간 '07~'09		2008 생명표	관찰기간 '08		관찰기간 '07~'09	
		G13	Beers	G13	Beers		G13	Beers	G13	Beers
전국	0.00785	0.00411	0.00672	0.00199	0.00621	0.00602	0.00257	0.00351	0.00187	0.00349
서울	0.01074	0.00434	0.00792	0.00275	0.00661	0.01001	0.00345	0.00379	0.00400	0.00407
부산	0.02586	0.00898	0.01124	0.00407	0.00842	0.01110	0.00586	0.00595	0.00418	0.00517
대구	0.03230	0.01184	0.01132	0.00724	0.00769	0.01928	0.00829	0.00456	0.00519	0.00598
인천	0.03420	0.01197	0.00757	0.00596	0.00712	0.01642	0.00747	0.00661	0.00355	0.00527
광주	0.04573	0.01543	0.01116	0.01321	0.00704	0.02898	0.01341	0.01327	0.00702	0.00645
대전	0.03951	0.01399	0.01213	0.00770	0.00608	0.02671	0.00740	0.00635	0.00392	0.00428
울산	0.06065	0.03203	0.01600	0.01249	0.00788	0.02988	0.00980	0.00961	0.00745	0.00581
경기	0.01581	0.00795	0.00717	0.00349	0.00676	0.00748	0.00269	0.00334	0.00221	0.00345
강원	0.03404	0.01434	0.01248	0.00977	0.01005	0.01991	0.00787	0.00730	0.00455	0.00537
충북	0.03038	0.01248	0.01241	0.00502	0.00917	0.02875	0.00945	0.00569	0.00529	0.00448
충남	0.02650	0.00857	0.00950	0.00595	0.00941	0.01639	0.00704	0.00503	0.00432	0.00453
전북	0.02131	0.00646	0.00889	0.00535	0.00797	0.01345	0.00473	0.00478	0.00347	0.00521
전남	0.02652	0.01632	0.01130	0.00630	0.00881	0.01325	0.00462	0.00602	0.00312	0.00469
경북	0.02536	0.00789	0.01139	0.00789	0.00942	0.01054	0.00391	0.00479	0.00264	0.00445
경남	0.02316	0.00999	0.01076	0.00808	0.01026	0.01589	0.00655	0.00668	0.00459	0.00579
제주	0.07409	0.03637	0.01898	0.00909	0.01481	0.02350	0.00884	0.00816	0.00763	0.00731
시도 평균	0.03288	0.01368	0.01126	0.00715	0.00859	0.01822	0.00696	0.00637	0.00457	0.00514





[그림 2-4] 사망확률 적합도와 평활도

2. 사망원인 생명표

가. 사망원인 생명표

사망원인에 대한 생명표는 주어진 사망구조에서 특정사인에 대한 연령별 위험률과 특정사인을 제거하였을 경우 신장(extension)되는 기대수명에 대한 정보를 제공한다. 생명표는 이탈을 일으키는 원인의 수에 따라서 하나인 경우를 단일생명표(single decrement table), 여러 가지인 경우를 다중생명표(multiple decrement table)로 구분할 수 있다. 일반 생명표의 경우 사망이라는 한 가지 요인으로 인하여 이탈이 발생하므로 단일생명표이며, 여러 가지 사인으로 인하여 이탈이 발생하는 사망원인 생명표는 다중생명표로 볼 수



있다. 사망원인과 관련해서 작성되는 생명표는 각 사인에 대한 연령별 위험률을 나타내는 다중생명표와 특정사인 i 가 제거된 경우에 대한 사망원인 생명표가 있다. 본 절의 내용은 미국 NCHS(1999)와 통계청 작성방법, 통계청(1996) 그리고 Samuel H. Preston et al(2001)을 참조하여 정리한 것임을 밝혀둔다.

전체 사인이 결합된 생명표에서 x 세와 $x+n$ 세 사이에 관측된 사망자수와 사인 i 에 의한 사망자수를 각각 ${}_nD_x$, ${}_nD_x(i)$ 라 하면 특정사인 i 에 의한 사망자수 ${}_nd_x(i)$ 와 x 세와 $x+n$ 세의 사인 i 비율은 다음과 같이 계산할 수 있다.

$$\begin{aligned} {}_nd_x(i) &= {}_nr_x^i \cdot {}_nD_x \\ {}_nr_x^i &= {}_nD_x(i) / {}_nD_x \end{aligned} \quad (4)$$

사인 비율 ${}_nr_x^i$ 를 이용하여 다음과 같은 추정량들을 정의할 수 있다.

$${}_nq_x^i = x \text{세와 } x+n \text{세의 사인 } i \text{로 인한 사망확률} = {}_nq_x \times {}_nr_x^i = {}_nd_x(i) / l_x$$

$${}_nq_x^{-i} = x \text{세와 } x+n \text{세의 사인 } i \text{를 제외한 다른 사인으로 인한 사망확률}$$

$$= {}_nq_x \times (1 - {}_nr_x^i) = \frac{{}_nd_x - {}_nd_x(i)}{l_x}$$

$${}_nd_x(i) = x \text{세와 } x+n \text{세의 사인 } i \text{로 인한 기대 사망자수} = {}_nq_x^i \times l_x$$

$${}_nd_x(-i) = x \text{세와 } x+n \text{세의 사인 } i \text{를 제외한 다른 사인으로 인한 기대 사망자수}$$

$$= {}_nq_x^{-i} \times l_x$$

여기서 특정사인을 제외한 사망확률 ${}_nq_x^{-i}$ 는 x 세인 사람이 사인 i 로 인한 사망을 제외하였을 때, 다른 사인에 의해 n 년 이내에 사망할 확률을 의미하는 것으로 특정사인을 제거한 사망확률과 구분할 필요가 있다. 사망확률 ${}_nq_x^{-i}$ 를 이용하여 작성되는 생명표의 기대여명은 주어진 사망구조에서 사망원인 i 를 제외하였을 때 증가되는 연령 x 의 기대여명을 의미하게 된다.

특정사인 i 로 인해 궁극적으로 사망하게 되는 연령 x 인 사람 수 l_x^i 와 그 비율 $R_x(i)$

는 다음과 같다. 여기서 $R_x(i)$ 는 연령 x 인 사람이 궁극적으로 사망원인 i 로 사망하게 될 확률로 위험률을 나타낸다.

$$l_x^i = \sum_{t=x}^w {}_n d_t(i), \quad (5)$$

$$R_x(i) = \sum_{t=x}^w \frac{l_x^i}{l_x}. \quad (6)$$

따라서 사망원인에 대한 다중생명표(multiple decrement table)는 ${}_n D_x$, ${}_n D_x(i)$, l_x , ${}_n q_x$, ${}_n q_x^i$, ${}_n d_x(i)$, l_x^i , $R_x(i)$ 로 구성되어 작성된다.

다중생명표와 달리 특정사인을 제거한 경우의 사망구조 변화를 파악하고자 할 때 작성되는 생명표가 사망원인 제거시 생명표(Cause-elimination Life Tables for Causes)이다. 사망원인 제거시 생명표를 작성하기 위해서는 사인 i 를 제거한 경우에 x -세인 사람이 n 년 까지 생존할 확률 ${}_n p_x^{(-i)}$ 이 필요하다. 이에 대해 여러 가지 접근 방법이 있으나, Chiang(1968)이 제안한 방법을 이용하여 소개하기로 한다. 모든 사망원인이 결합된 사망력(Force of Mortality)은 x 세 까지 생존한 사람이 순간적으로 사망할 확률을 의미하는 것으로 다음과 같이 정의된다.

$$\mu_x = \lim_{n \rightarrow 0} {}_n m_x$$

여기서 ${}_n m_x$ 는 연령 x 와 $x+n$ 사이의 사망률을 나타낸다. 사망원인 i 에 의한 사망률을 ${}_n m_x^i$ 라고 할 때, 사망력 μ_x^i 는 다음과 같이 정의된다.

$$\mu_x^i = \lim_{n \rightarrow 0} {}_n m_x^i.$$

사망원인이 K 개인 경우, $\sum_{i=1}^K {}_n m_x^i = {}_n m_x$ 이므로 $\mu_x = \sum_{i=1}^K \mu_x^i$ 가 성립하며, 생존확률 ${}_n p_x$ 와 사망력 μ_x 간에는 다음과 같은 관계식이 성립한다.



$${}_n p_x = \exp\left(-\int_n^{x+n} \mu_y dy\right). \quad (7)$$

식(7)은 다음과 같이 표현할 수 있다.

$${}_n p_x = \exp\left(-\int_n^{x+n} \mu_y^1 dy\right) \cdot \cdots \cdot \exp\left(-\int_n^{x+n} \mu_y^K dy\right) = {}_n p_x^1 \cdots {}_n p_x^K \quad (8)$$

여기서 ${}_n p_x^i = \exp\left(-\int_n^{x+n} \mu_y^i dy\right)$ 이며, 사망원인 i 에 대한 x 세인 사람의 n 년까지 생존 확률을 나타낸다. Chiang(1968)은 사인 i 에 의한 사망력이 전체 사망원인에 대한 사망력에 비례한다는 다음과 같은 가정을 하였다.

$$\mu_y^i = {}_n r_x^i \cdot \mu_y, \quad x \leq y \leq x+n.$$

따라서 ${}_n p_x^{(i)}$ 와 ${}_n p_x$ 사이에 다음의 관계식이 성립한다.

$$\begin{aligned} {}_n p_x^i &= \exp\left(-\int_n^{x+n} \mu_y^i dy\right) = \exp\left(-{}_n r_x^i \cdot \int_n^{x+n} \mu_y dy\right) \\ &= \left[\exp\left(-\int_n^{x+n} \mu_y dy\right)\right]^{r_x^i} = ({}_n p_x)^{r_x^i}. \end{aligned} \quad (9)$$

식(8)과 (9)을 이용하여 사인 i 가 제거된 경우 x 세인 사람의 n 년 까지 생존 확률 ${}_n p_x^{(-i)}$ 과 사망확률 ${}_n q_x^{(-i)}$ 는 각각 다음과 같이 구해진다.

$${}_n p_x^{(-i)} = \frac{{}_n p_x}{{}_n p_x^{n r_x^i}} = {}_n p_x^{(1 - n r_x^i)}, \quad (10)$$

$${}_n q_x^{(-i)} = 1 - {}_n p_x^{(1 - n r_x^i)}. \quad (11)$$

식(11)의 사망확률에 대해 일반생명표와 같은 방법으로 사망자수 ${}_n d_x^{(-i)}$, 생존자수 $l_x^{(-i)}$, 정지인구 ${}_n L_x^{(-i)}$, 총 생존연수 $T_x^{(-i)}$, 사망원인 i 제거시 연령 x 의 기대여명 $e_x^{(-i)}$ 를 산출하며, 전체사인을 결합한 기대여명 e_x 과 $e_x^{(-i)}$ 의 차이가 특정사인이 제거되었을 때 증가된 기대여명이 된다. 각각의 산출단계를 정리하면 다음과 같다.

(특정사인 제거시 사망자수) ${}_n d_x^{(-i)} = {}_n q_x^{(-i)} \times l_x$

(특정사인 제거시 생존자수) $l_{x+n}^{(-i)} = l_x^{(-i)} - {}_n d_x^{(-i)}$

(특정사인 제거시 정지인구) ${}_n L_x^{(-i)} = n \times \left(\frac{l_x^{(-i)} + l_{x+n}^{(-i)}}{2} \right)$

(0세) ${}_1 L_0^{(-i)} = f_0 l_0^{(-i)} + (1 - f_0) l_1^{(-i)}$, f_0 는 0세 분리계수

(1~4세) ${}_4 L_1^{(-i)} = f_{1-4} l_1^{(-i)} + (1 - f_{1-4}) \cdot l_5^{(-i)}$, f_{1-4} 는 1~4세 분리계수

(상한 연령) ${}_{\infty} L_w^{(-i)} = \frac{l_w^{(-i)} \cdot e_w}{1 - {}_{\infty} R_w(i)}$, e_w =일반생명표 최고연령의 기대여명

(특정사인 제거시 총 생존연수) $T_x^{(-i)} = T_{x+n}^{(-i)} + {}_n L_x^{(-i)}$

(특정사인 제거시 기대여명) $e_x^{(-i)} = \frac{T_x^{(-i)}}{l_x^{(-i)}}$

(특정사인 제거시 증가된 기대여명) $e_x - e_x^{(-i)}$



현재 통계청에서는 다중생명표와 사인 제거시 생명표를 모두 포함하여 사망원인 생명표를 작성하고 있다. 작성방법은 특정사인 i 에 대해 식(4)를 이용하여 ${}_n r_x^i$ 를 계산하고 사인으로 인한 사망자수 ${}_n d_x(i)$ 와 사인을 제외한 경우에 대한 사망자수 ${}_n d_x(-i)$ 를 아래의 식을 적용하여 산출한다. 이후 일반적인 생명표와 동일한 방법으로 작성되며 자세한 내용은 2절에 수록하였다.

$${}_n d_x(i) = {}_n d_x \times {}_n r_x^i,$$

$${}_n d_x(-i) = {}_n d_x - {}_n d_x(i).$$

나. 실증 분석

2008년 남자의 특정사인이 악성신생물인 경우에 대하여 다중생명표와 사인을 제외한 후 사망원인 생명표(사망확률 ${}_n q_x^{-i}$ 을 이용), 사인을 제거한 후 사망원인 생명표(사망확률 ${}_n q_x^{(-i)}$ 을 이용), 통계청 사망원인 생명표를 작성하고 각각 <표 2-11> ~ <표 2-14>에 수록하였다. <표 2-11> 다중생명표는 악성신생물이라는 사인으로 인해 연령별 궁극적으로 사망할 위험률 $R_x(i)$ 와 사인 i 로 인한 연령별 사망확률 ${}_n q_x^i$ 등이 산출되어 있다. 50~54세에서 악성신생물로 사망할 확률은 0.008688이며 50~54세인 사람이 악성신생물로 인해 궁극적으로 사망할 확률(위험률)은 0.290462이다. 악성신생물에 의한 위험률은 영아 0.284513부터 연령이 증가함에 따라 꾸준히 증가하다가 50~54세 연령구간을 기점으로 감소하는 추세를 보이고 있으며, 85세 이상에서 악성신생물로 사망할 확률 혹은 위험률은 0.159727임을 알 수 있다. <표 2-12>의 악성신생물을 제외하고 구한 생명표는 각 연령별 사망확률이 사인비율 만큼 감소하게 되며 85세 이상 사망확률 ${}_n q_x^{-i}$ 의 경우 0.840273으로 약 0.16 정도 감소하며, 이 경우 기대수명은 80.60으로 4.06세가 증가하게 되는 것으로 나타났다. <표 2-13>의 악성신생물 제거시 생명표는 현재의 사망구조에서 악성신생물이 제거된 후의 사망구조를 보여주며, 기대수명은 80.36으로 3.82세가 증가하게 된다. <표 2-14>는 통계청의 작성방법을 이용하여 작성된 악성신생물에 대한 사망원인 생명표로 기대수명은 81.52세로 4.98세가 증가되는 것으로 나타났다. 사망원인 생명표 작성방법에 따라서 생명표를 작성한 결과를 살펴보면, 방법에 따라 기대수명에 차이가 발생하며, 그로 인해 증가되는 기대수명에도 차이가 있으나, 모두 특정사인으로 인해

변화되는 사망구조와 기대수명에 대한 정보를 제공하는 것으로 볼 수 있다. 그러나 사망 원인 생명표는 특정사인이 제거된 경우의 사망구조 변화를 알아보려고 작성되는 것이므로 보다 엄밀한 의미에서 볼 때, 식(11)의 사망원인 제거시 사망확률을 이용하여 작성되는 것이 바람직할 것으로 판단된다.

〈표 2-11〉 2008 남자 : 다중생명표(악성신생물)

연령	${}_nD_x$	${}_nD_x(i)$	${}_nr_x^i$	l_x	q_x	q_x^i	$d_x(i)$	역누적	$R_x(i)$
0	842	8	0.00950	100000	0.003970	0.000038	4	28451	0.284513
1	267	31	0.11610	99603	0.001138	0.000132	13	28448	0.285610
5	216	44	0.20370	99490	0.000718	0.000146	15	28434	0.285803
10	250	54	0.21600	99418	0.000701	0.000151	15	28420	0.285862
15	690	59	0.08551	99348	0.001937	0.000166	16	28405	0.285911
20	958	87	0.09081	99156	0.002846	0.000258	26	28388	0.286300
25	1553	134	0.08628	98874	0.003727	0.000322	32	28363	0.286858
30	1652	229	0.13862	98505	0.004113	0.000570	56	28331	0.287608
35	3111	506	0.16265	98100	0.006656	0.001083	106	28275	0.288223
40	4852	1014	0.20899	97447	0.011012	0.002301	224	28169	0.289065
45	8284	2175	0.26255	96374	0.018127	0.004759	459	27944	0.289956
50	9702	3143	0.32395	94627	0.026818	0.008688	822	27486	0.290462
55	10174	3976	0.39080	92090	0.039748	0.015534	1430	26664	0.289539
60	11304	4831	0.42737	88429	0.056833	0.024289	2148	25233	0.285348
65	17379	7434	0.42776	83403	0.092669	0.039640	3306	23085	0.276790
70	19985	7941	0.39735	75675	0.152998	0.060793	4601	19779	0.261371
75	17348	6153	0.35468	64096	0.244230	0.086624	5552	15179	0.236809
80	14314	3721	0.25996	48442	0.389022	0.101128	4899	9626	0.198718
85+	14049	2244	0.15973	29597	1.000000	0.159727	4727	4727	0.159727

〈표 2-12〉 2008 남자: 사망원인(악성신생물)을 제외한 사인생명표

연령	q_x	l_x	${}_nq_x^{-i}$	l_x^{-i}	${}_nd_x(-i)$	L_x^{-i}	T_x^{-i}	e_x^{-i}	e_x
0	0.003970	100000	0.003932	100000	393	99680	8059629	80.60	76.54
1	0.001138	99603	0.001006	99607	100	398193	7959949	79.91	75.84
5	0.000718	99490	0.000572	99507	57	497390	7561756	75.99	71.93
10	0.000701	99418	0.000550	99450	55	497111	7064366	71.03	66.98
15	0.001937	99348	0.001771	99395	176	496535	6567254	66.07	62.02
20	0.002846	99156	0.002588	99219	257	495453	6070720	61.19	57.14
25	0.003727	98874	0.003405	98962	337	493968	5575267	56.34	52.29
30	0.004113	98505	0.003543	98625	349	492252	5081299	51.52	47.48
35	0.006656	98100	0.005573	98276	548	490010	4589046	46.70	42.67
40	0.011012	97447	0.008711	97728	851	486512	4099037	41.94	37.93
45	0.018127	96374	0.013367	96877	1295	481146	3612525	37.29	33.33
50	0.026818	94627	0.018130	95582	1733	473577	3131378	32.76	28.89
55	0.039748	92090	0.024215	93849	2273	463563	2657802	28.32	24.62
60	0.056833	88429	0.032544	91576	2980	450431	2194238	23.96	20.53
65	0.092669	83403	0.053029	88596	4698	431235	1743807	19.68	16.60
70	0.152998	75675	0.092204	83898	7736	400150	1312572	15.64	13.03
75	0.244230	64096	0.157606	76162	12004	350802	912422	11.98	9.90
80	0.389022	48442	0.287894	64159	18471	274616	561620	8.75	7.26
85+	1.000000	29597	0.840273	45688	38390	287005	287005	6.28	5.28



〈표 2-13〉 2008 남자: 사망원인(악성신생물)을 제거한 사인생명표

연령	q_x	l_x	${}_nq_x^{(-i)}$	$l_x^{(-i)}$	${}_nd_x(-i)$	$L_x^{(-i)}$	$T_x^{(-i)}$	$e_x^{(-i)}$	e_x
0	0.003970	100000	0.003933	100000	393	99680	8036115	80.36	76.54
1	0.001138	99603	0.001006	99607	100	398193	7936436	79.68	75.84
5	0.000718	99490	0.000572	99507	57	497390	7538243	75.76	71.93
10	0.000701	99418	0.000550	99450	55	497111	7040852	70.80	66.98
15	0.001937	99348	0.001771	99395	176	496535	6543741	65.84	62.02
20	0.002846	99156	0.002588	99219	257	495452	6047207	60.95	57.14
25	0.003727	98874	0.003406	98962	337	493968	5551754	56.10	52.29
30	0.004113	98505	0.003544	98625	350	492252	5057786	51.28	47.48
35	0.006656	98100	0.005576	98276	548	490008	4565535	46.46	42.67
40	0.011012	97447	0.008721	97728	852	486507	4075527	41.70	37.93
45	0.018127	96374	0.013399	96875	1298	481131	3589020	37.05	33.33
50	0.026818	94627	0.018210	95577	1740	473535	3107889	32.52	28.89
55	0.039748	92090	0.024406	93837	2290	463458	2634354	28.07	24.62
60	0.056833	88429	0.032950	91547	3017	450192	2170896	23.71	20.53
65	0.092669	83403	0.054129	88530	4792	430670	1720704	19.44	16.60
70	0.152998	75675	0.095227	83738	7974	398755	1290034	15.41	13.03
75	0.244230	64096	0.165315	75764	12525	347507	891279	11.76	9.90
80	0.389022	48442	0.305537	63239	19322	267890	543772	8.60	7.26
85+	1.000000	29597	1.000000	43917	43917	275882	275882	6.28	5.28

〈표 2-14〉 2008 남자: 통계청 사망원인(악성신생물) 생명표

연령	q_x	l_x	${}_nd_x(-i)$	$l_x(-i)$	$L_x(-i)$	$T_x(-i)$	$e_x(-i)$	${}_nq_x(-i)$	e_x
0	0.003970	100000	393	100000	99680	8151949	81.52	0.003932	76.54
1	0.001138	99603	100	99607	398193	8052270	80.84	0.001006	75.84
5	0.000718	99490	57	99507	497390	7654077	76.92	0.000572	71.93
10	0.000701	99418	55	99450	497112	7156686	71.96	0.000550	66.98
15	0.001937	99348	176	99395	496535	6659575	67.00	0.001770	62.02
20	0.002846	99156	257	99219	495454	6163040	62.12	0.002586	57.14
25	0.003727	98874	337	98962	493970	5667586	57.27	0.003402	52.29
30	0.004113	98505	349	98626	492256	5173616	52.46	0.003539	47.48
35	0.006656	98100	547	98277	490017	4681359	47.63	0.005563	42.67
40	0.011012	97447	849	97730	486528	4191342	42.89	0.008685	37.93
45	0.018127	96374	1288	96881	481185	3704814	38.24	0.013297	33.33
50	0.026818	94627	1716	95593	473676	3223629	33.72	0.017947	28.89
55	0.039748	92090	2230	93877	463812	2749953	29.29	0.023754	24.62
60	0.056833	88429	2878	91647	451042	2286141	24.94	0.031401	20.53
65	0.092669	83403	4423	88770	432791	1835099	20.67	0.049824	16.60
70	0.152998	75675	6978	84347	404290	1402308	16.63	0.082724	13.03
75	0.244230	64096	10102	77369	361591	998018	12.90	0.130569	9.90
80	0.389022	48442	13946	67267	301471	636427	9.46	0.207326	7.26
85	1.000000	29597	24870	53321	334956	334956	6.28	0.466414	5.28



제5절 결 론

1. 요약

생명표는 주어진 인구집단을 대상으로 일정기간을 관찰하여 연령에 따라 변화하는 사망 상태를 보여주는 통계표로 사망확률, 사망자수, 기대여명 등으로 구성된다. 생명표 작성은 관찰 기간에 따른 사망자료를 기반으로 기초 사망률을 산출하고 보정을 통해 사망확률을 추정한 후 일반적인 일련의 과정에 의해 이루어지게 된다. 사망확률 추정은 적합도(fitness)와 평활도(smoothness) 측면을 고려하여 기초 사망확률을 보정하는 것으로 생명표 작성을 위한 기초작업이라고 할 수 있다. 본 연구는 지역 단위로 작성되는 시도 생명표와 특정 사망원인이 제거되었을 때를 가정하여 작성되는 사망원인 생명표의 사망확률 추정방법에 대한 이론 및 실증분석을 통하여 추정방법에 따른 결과를 비교하고 적절한 추정방안을 찾고자 하는데 목적이 있다.

우리나라 시도 생명표는 2005년을 시작으로 3년 주기로 16개 시도에 대해 간이생명표로 작성되며, 기초자료로는 작성되는 해의 단년 자료를 이용하고 있다. 이러한 방법으로 작성되는 시도 생명표는 시의성과 전국 단위로 작성되는 일반생명표와 비교가 가능하다는 측면에 의미가 있다고 할 수 있다. 그러나 생명표 작성에 필요한 기초자료 측면에서 보면 관찰기간이 충분하지 못하다는 점과 그로 인해 추정되는 각 세 사망확률의 불안정을 초래할 가능성이 있다. 미국, 캐나다 등 해외사례를 살펴보았을 때, 지역 단위로 작성되는 생명표는 센서스를 실시하는 해를 중심으로 3년 자료를 이용하여 작성되고 있으며, 전국 단위와 비교하기 위하여 동일한 방법으로 일반생명표를 작성하고 있음을 알 수 있다. 본 연구에서는 2008년 작성된 시도 생명표에 대해 관찰기간을 2007~2009년으로 3년을 고려하여 16개 시도에 대해 성별로 나누어 작성하고 비교하였다.

시도별 각 세 사망확률의 평활도를 높이기 위하여 보정방법으로 Beers 방법과 Greville 13차항 방법을 적용하였다. 보정방법과 관찰기간을 고려하여 4가지 방법을 적용하고 2008년 시도별 사망확률과 적합도와 평활도 그리고 기대수명을 비교하였다. 보정된 사망확률 그림을 통해 3년 자료를 이용한 경우가 1년 자료를 이용하였을 때보다 안정적인 사망확률을 산출하였으며, Greville 13차항 방법과 Beers 방법은 저연령 사망확률의 최저점 형태에 약간의 차이가 나타났으나, 그 외 연령에서는 뚜렷한 차이를 보이지 않았다. 기대수명은 검토한 4가지 방법이 대체적으로 2008년 생명표 기대수명보다 낮았으며, 1년 보다는 3년 자료를 사용하였을 때, Greville 방법보다는 Beers 방법이 다소 낮게 산출되는 경향을 보였다. 추정된 사망확률의 적합도에서 2008년 생명표의 경우 전국과 시도의 적합도가 큰 차이를 보이거나 검토한 방법에서는 전국과 시도간 적합도의 차이가 다소 줄어드는 것을 볼 수 있다. 평활도를 살펴보면 2008년 생명표가 가장 낮으며 적



합도와 마찬가지로 전국과 시도간 차이가 큰 것으로 나타났다. 검토한 방법들의 평활도는 2008년 생명표와 비교하였을 때 모두 높게 나타났으며, 울산, 제주도와 같은 지역을 고려한다면 관찰기간을 3년으로 하는 것이 적절한 것으로 보여진다.

사망원인 생명표는 특정사인으로 사망하게 될 확률(위험률)과 특정사인이 제거되었을 때 변화되는 사망구조와 그에 따라 증가하는 기대여명에 대한 정보를 제공하는 생명표이다. 본 연구에서는 사망원인 생명표 작성에 필요한 기본적인 개념과 특정사인이 제거되었을 때의 사망확률 추정방법에 대하여 사망원인 제외 시 사망확률과 구분하여 이론적 배경을 살펴보았다. 2008년 남자에 대해 특정사인이 악성신생물인 경우에 대하여 실증분석을 실시한 결과 각 방법별로 기대수명이 81.52세, 80.60세, 80.36세로 추정되었다. 연구 결과로부터 사망원인으로 인해 변화된 사망확률 추정 결과에 있어서 차이가 발생하며 기대수명에도 영향을 미치고 있음을 알 수 있다. 3가지 방법이 특정 사망원인으로 인한 사망상태 변화에 대한 정보를 제공하고 있으나 보다 엄밀한 의미에서 볼 때, 식(11)에 의한 특정사인 제거시 사망확률을 사용하여 사망원인 생명표를 작성하는 것이 바람직할 것으로 보인다.

2. 결론 및 제언

생명표는 주어진 인구집단에 대한 사망구조와 기대여명 등을 나타내는 중요한 사회지표이며, 생명표 작성의 기초단계는 연령별 사망확률 추정에 있다고 할 수 있다. 본 연구는 시도 생명표와 사망원인 생명표 작성에 필요한 사망확률 추정방법에 대하여 검토하였으며, 연구결과를 정리하면 다음과 같다. (1)시도 생명표의 각 세 사망확률 안정성 확보를 위해서는 현재의 관찰기간 1년은 충분하지 않으며, 3년 정도의 기간이 필요하다 (2)일반생명표 사망확률에 사용되고 있는 Greville 9차항 방법 외 13차항 방법과 Beers 방법을 적용한 결과 기대수명은 다소 낮게 추정되었으나 적합도와 평활도 측면에서 효율적인 것으로 나타났다 (3)3년 자료에 대해 Greville 방법과 Beers 방법으로 보정하였을 때, 1년 자료를 이용하여 시도 생명표의 각 세 사망확률 추정 시 발생 가능한 불안정성을 어느 정도 해소할 수 있는 것으로 판단된다 (4)사망원인 생명표의 사망확률은 현재의 추정 방법도 특정사인에 의한 사망확률에 따른 변화에 대한 정보를 가지고 있으나 특정사인 제거시 사망확률을 이용하는 것이 보다 엄밀한 의미에서 타당하다고 판단된다.

시도 생명표는 전국 사망수준과 각 16개 시도의 사망수준 비교와 시의성을 담보하기 위하여 작성되는 해의 단년치 자료를 이용하여 작성되어 왔으나 시의성 못지않게 안정성 또한 중요하게 고려되어야 할 것이다. 사망원인 생명표 또한 예방과 치료에 필요한 의료정책 수립에 대한 기초자료로 그 의미가 있으며, 보다 타당한 방법을 적용하여 작성될 필요가 있다. 2011년 시도 생명표와 사망원인 생명표 작성이 다가오는 시점에서 본 연구결과가 생명표 품질 개선에 도움이 되기를 기대한다.

참고문헌

- 구자홍 (2009), 「인구통계학의 이론과 실제」, 교우사.
- 박유성, 박성근, 최보승, 김기환 (2009), Beers와 Greville 보정방법에 의한 사망확률 보정방법 비교 연구, *Journal of the Korean Data Analysis Society*, 11, 1, 97-110.
- 통계청 (1996), 사망원인 생명표, 인구동태 학술연구 용역보고서.
- 통계청 (2010), 2009년 생명표.
- Chiang, C.L. (1968), 「An Introduction to Stochastic Processes in Biostatistics」, New York: Wiley
- Coale, A and Kisker, E. (1990), Defects in data on old age mortality in the United States : new procedures for calculating approximately accurate mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, 4, 1-31.
- London, R. L. (1981), In Defense of Minimum-Ro Linear Compound Graduation, and a Simple Modification for its Improvement, *ARCH, Chicago: Society Acturies*, pp. 75-88.
- National Center for Health Statistics (1998), U.S. decennial life tables for 1989-91, vol 1, no. 2, methodology of the national and State life tables.
- National Center for Health Statistics (1999), Method for constructing complete annual U.S. life tables, *Vital Health Stat* 2(129).
- National Center for Health Statistics (1999), U.S. decennial life tables for 1989-91, vol 1, no. 4, United States life tables eliminating certain causes of death.
- Samuel H. Preston, Patrick Heuveline, Michel Guillot(2001), *Demography*, Blackwell
- Statistics Canada(2006), *Life Tables, Canada, Provinces and Territories 2000 to 2002*.