



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이 학 석 사 학 위 논 문

프레일티 모형을 이용한 주요 암 종별
부산시 소지역별 생존율 비교 연구



2019년 2월

부 경 대 학 교 대 학 원

통 계 학 과

최 희 원

이 학 석 사 학 위 논 문

프레일티 모형을 이용한 주요 암 종별
부산시 소지역별 생존율 비교 연구

지도교수 노 맹 석
공동지도교수 하 일 도

이 논문을 석사 학위논문으로 제출함.

2019년 2월

부 경 대 학 교 대 학 원

통 계 학 과

최 희 원

최희원의 이학석사 학위논문을 인준함.

2019년 2월 22일



위원장	이학박사	장대홍 (인)
위원	이학박사	하일도 (인)
위원	이학박사	노맹석 (인)

목 차

표 차례	ii
그림 차례	v
I 서론	1
II 연구방법	3
2.1 카플란 마이어 추정량	3
2.2 프레일티 모형	4
2.3 가능도 추론	5
2.4 자료설명	7
2.5 분석방법	9
III 결과	11
3.1 생존율 추정	11
3.2 프레일티 모형적합 및 시각화 결과	19
IV 요약 및 결론	36
참고문헌	37
부록 A.1 SAS 코드	40
A.2 R 코드	43
B 추가 그림	45

표 차례

표 1. 암 등록자료 분석 변수 설명	9
표 2. 성별에 따른 연령별 및 병기별 위암 생존율	14
표 3. 성별에 따른 연령별 및 병기별 대장암 생존율	15
표 4. 성별에 따른 연령별 및 병기별 폐암 생존율	16
표 5. 성별에 따른 연령별 및 병기별 간암 생존율	17
표 6. 위암: 프레일티 모형 적합 결과	20
표 7. 대장암: 프레일티 모형 적합 결과	20
표 8. 폐암: 프레일티 모형 적합 결과	21
표 9. 간암: 프레일티 모형 적합 결과	21
표 10. 위암: 프레일티 모형 적합 결과: 박탈지수 포함	31
표 11. 대장암: 프레일티 모형 적합 결과: 박탈지수 포함	32
표 12. 폐암: 프레일티 모형 적합 결과: 박탈지수 포함	32
표 13. 간암: 프레일티 모형 적합 결과: 박탈지수 포함	33

그림 차례

그림 1. 성별 및 연령에 따른 병기별 위암 카플란 마이어 생존곡선	18
그림 2. 동별 위암의 로그 프레일티에 대한 95% 신뢰구간	23
그림 3. 동별 대장암의 로그 프레일티에 대한 95% 신뢰구간	23
그림 4. 동별 폐암의 로그 프레일티에 대한 95% 신뢰구간	24
그림 5. 동별 간암의 로그 프레일티에 대한 95% 신뢰구간	24
그림 6. 생존시간에 따른 위암 프레일티 효과: 서동	25
그림 7. 생존시간에 따른 대장암 프레일티 효과: 감천동	26
그림 8. 생존시간에 따른 폐암 프레일티 효과: 반송동	26
그림 9. 생존시간에 따른 간암 프레일티 효과: 다대동	27
그림 10. 부산시 위암 질병지도 (9분위)	28
그림 11. 부산시 대장암 질병지도 (9분위)	28
그림 12. 부산시 폐암 질병지도 (9분위)	29
그림 13. 부산시 간암 질병지도 (9분위)	29
그림 14. 박탈지수에 따른 로그 프레일티 추정값	30
그림 15. 동별 위암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함 ..	34
그림 16. 동별 대장암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함	34
그림 17. 동별 폐암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함 ..	35
그림 18. 동별 간암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함 ..	35
그림 B.1. 성별 및 연령에 따른 병기별 대장암 카플란 마이어 생존곡선	45
그림 B.2. 성별 및 연령에 따른 병기별 폐암 카플란 마이어 생존곡선	46
그림 B.3. 성별 및 연령에 따른 병기별 간암 카플란 마이어 생존곡선	47

A comparative study on survival rates of Busan small area
by major type of cancer using frailty models

Hee Won Choi

Department of Statistics, The Graduate School,
Pukyong National University

Abstract

Cancers, also called malignant neoplasm, correspond to International Classification of Diseases (ICD-10) C00-C97. They are still one of the major causes of death in Korea and mortality rates have been increasing until recently. In particular, the cancer mortality rates of Busan have been higher than those of the whole country or Seoul since 2001, and the gap became more and more widening.

Using cancer registry data of Busan, in this paper, we conducted the survival analysis for the survival time which is defined as time from initial diagnosis to death of cancer. For this purpose, we compared the survival rates estimated by gender, age groups, and summary stages of stomach cancer, colorectum cancer, lung cancer and liver cancer classified as major cancer types in Korea. As the survival rates of cancers may be different according to small area, the analysis was performed using the frailty survival models considering the unit of Eup-Myeon-Dong as random effects. Also, various and effective visualizations for the fitting results was presented including the disease mapping of Busan. In addition, the inference results were compared according to whether or not the deprivation index variable indicating the socioeconomic index by small area was included. As a result, the deprivation index reduced the significance of risk for cancer death among dongs.

keywords: Cancer, frailty model, random effect, survival analysis

I 서론

국제질병분류코드 ICD-10의 C00-C97에 해당하는 악성신생물, 즉, 암은 여전히 우리나라에서 주요한 사망원인 중 하나이다. 일반적으로 성별, 연도별 또는 지역별로 인구의 연령구조가 다르므로 발생률 및 사망률에 대한 보다 정확한 비교를 위해 동일한 기준으로 표준인구를 활용하여 차이를 보정한 연령표준화율 (age-standardized rate; ASR)을 사용한다. 우리나라 암 발생현황을 살펴보았을 때, 암등록통계에서 2000년 주민등록인구를 표준인구로 사용하여 표준화를 시행한 결과, 2012년 이후 4년 간 연령표준화 암 발생률이 약 6.1%씩 감소하는 추세를 보였다 (보건복지부, 2015). 우리나라 암 사망의 측면에서 2000년 이래로 암 사망률이 지속적으로 증가해왔으며 2017년 기준 사망원인 1위에 해당하였다 (통계청 사망원인통계, 2017). 특히, 연령표준화 암 사망률에서 최근 20년 동안 부산광역시가 전국 또는 서울특별시의 수준보다 높았으며 그 차이가 크게 줄어들지 않고 있다. 성별로 살펴보았을 때 비슷한 경향을 보였으며, 주요 암 종별로는 위암, 대장암은 전체 암과 비슷하게 전국보다 표준화발생률이 더 높았으나 폐암, 간암의 경우 전국보다 더 낮은 수치를 보이는 연도가 존재하였다.

국내에서 시행된 암 관련 연구로는 암 수술 전후 신체에 영향을 미치는 예후인자에 대한 의학적 연구, 암 환자의 삶의 질에 대한 연구 등이 다수 존재하며 (Paik, 2002; Kim과 Tae, 2011), 암 생존율에 대해서도 영향요인을 다룬 연구를 시행하였다 (Kim, 2003; Lee와 Lee, 2017). 또한, 지역의 단위가 작을수록 건강격차가 증가한다는 연구결과에도 불구하고

(Yoon, 2010), 암 생존자료에 대한 소지역별 분석은 부족한 실정이다.

본 연구에서는 부산시 암 등록자료를 활용하여 우리나라에서 주요 암 종류로 분류되는 위암, 대장암, 폐암 및 간암 환자에 대하여 암 종별로 나타나는 생존율 추세가 다를 것이라 판단하고 암 종별 분석을 시행하였다. 먼저, 성별, 연령군, 및 병기에 따라 추정된 생존율 결과를 비교하였다. 구체적으로, 소지역에 따라 암 생존율이 다를 것이라 판단하여 변량효과(random effect)로 읍면동 단위를 고려한 프레일티 모형 (frailty survival model)을 이용하여 분석을 진행하였다. 여기서 프레일티란 관측할 수 없는 변량효과를 의미하며, 변량효과는 다변량 자료 사이의 의존성을 모형화하기에 유용하다. 또한, 모형적합 결과 효과적인 시각화를 위해 부산시 질병지도를 작성하였으며, 소지역별 사회·경제적 지수를 나타내는 지역 박탈지수 변수의 포함 여부에 따라 적합 결과를 비교하였다.

본 연구의 목적은 다음과 같다. 생존분석을 통해 각 특성에 따른 암 생존율 및 추세를 서로 비교하고, 부산광역시 암 사망환자의 소지역 간 이질성을 살펴보고자 하였다. 이러한 격차를 설명하기 위해 부산지역 암센터의 암 등록자료를 프레일티 생존모형에 적합하였다. 본 연구에서는 변량효과로서 105개로 통합한 읍면동에 대한 변수를 고려하여 모형에 적합하였다.

본 연구의 구성은 다음과 같다. 2장에서 카플란-마이어 추정량 (Kaplan-Meier estimator) 및 프레일티 모형에 대하여 간단히 기술하고 분석자료 및 분석방법에 대하여 소개한다. 생존율, 카플란-마이어 생존곡선 및 프레일티 모형 적합에 대한 결과 및 해석을 취합하여 3장에 제시하였다. 4장에서 전체적인 분석결과에 대해 논의하고 결론을 도출한다. 마지막으로 분석에 사용한 SAS 및 R 코드와 추가적인 분석결과가 부록에 주어진다.

II 연구방법

2.1 카플란 마이어 추정량

무정보적 (non-informative) 중도절단일 때 일반적으로 개인의 생존시간들은 서로 독립이며 중도절단은 생존시간에 무정보적, 즉, 중도절단 분포가 생존시간 분포로부터의 관심 있는 모수에 의존하지 않는다고 가정한다 (Klein과 Moeschberger, 2003). 이 때, 생존함수에 대한 추정량으로 가장 널리 알려져 있고 보편적으로 사용하는 비모수적 (non-parametric) 추정 방법 중 하나가 카플란 마이어 추정량이며 (Kaplan과 Meier, 1958), 누적 한계 추정량 (product-limit estimator) 이라고도 불린다. 이러한 가정을 바탕으로 T_i ($i=1, \dots, n$)를 i 번째 개체 (subject)의 생존시간 (survival time), 순서대로 배열된 관찰 생존시간을 $y_{(k)}$ ($k=1, \dots, D$), 각 시점에 사건이 발생한 수를 $d_{(k)}$, 위험에 노출된 관측치의 수를 $n_{(k)}$ 라고 할 때, K-M 추정량을 다음과 같이 조건부 확률의 곱으로 표현할 수 있다:

$$\hat{S}(t) = \prod_{k: y_{(k)} \leq t} \left\{ 1 - \frac{d_{(k)}}{n_{(k)}} \right\}$$

2.2 프레일티 모형

콕스 비례위험 모형 (Cox's proportional hazards model)의 한 확장형인 프레일티 모형은 관측이 불가능한 변량효과를 의미하는 프레일티를 고려한다 (Duchateau와 Janssen, 2008). i 번째 그룹 또는 집단 (cluster)의 j 번째 개체 또는 관측치 (observation)에 대한 생존시간을 T_{ij} ($i = 1, \dots, n, j = 1, \dots, n_i$)라 하고, 이에 해당하는 중도절단 시간 (censoring time)을 C_{ij} 로 정의하였을 때 관찰할 수 있는 확률변수 (random variable)와 중도절단 여부에 대한 표기는 각각 다음과 같다:

$$Y_{ij} = \min(T_{ij}, C_{ij}) \text{ and } \delta_{ij} = I(T_{ij} \leq C_{ij}),$$

여기서 $I(\cdot)$ 는 지시함수 (indicator function)이며, q 는 집단의 수, $n = \sum_i n_i$ 는 전체 관측치 수를 나타낸다. i 번째 집단에 대한 프레일티가 $U_i = u_i$ 로 주어졌을 때, T_{ij} 에 대한 조건부 위험함수는 다음과 같다:

$$\lambda_{ij}(t|u_i) = \lambda_0(t) \exp(x_{ij}^T \beta) u_i, \quad (2.1)$$

여기서 $\lambda_0(\cdot)$ 는 미지의 기저위험함수, $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ 는 공변량 (covariates) 벡터, $\beta = (\beta_1, \dots, \beta_p)^T$ 는 미지의 회귀모수 벡터를 의미한다. 일반적으로 프레일티에 대하여 로그정규 (log-normal) 분포 또는 감마 (Gamma) 분포를 사용하는 것으로 알려져 있다 (Ha 등, 2017), 본 연구에서는 로그 프레일티 (log-frailty)가 다음의 분포를 따르는 로그정규 프

레일티 모형을 가정하였다:

$$V_i = \log(U_i) \sim iid N(0, \sigma^2).$$

2.3 가능도 추론

모형 (2.1)에 대해 기술적으로 추론하기 위해 가능도 함수 (likelihood function)를 일반적으로 사용한다. 모형 (2.1)의 프레일티 모형에 대한 다단계 가능도 (hierarchical likelihood; h-likelihood)는 다음과 같이 표기한다 (Lee와 Nelder, 1996):

$$h = h(\beta, u, \lambda_0, \alpha) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}, \quad (2.2)$$

여기서 각 요소들은 프레일티 $U_i = u_i$ 가 주어졌을 때 y_{ij} 와 δ_{ij} 의 조건부 로 그 가능도 함수를 나타내는

$$\sum_{ij} \ell_{ij} = \sum_{ij} \ell_{1ij}(\beta, \lambda_0; \delta_{ij} | u_i) = \delta_{ij} \{ \log \lambda_0(y_{ij}) - \eta_{ij} \} - \{ \Lambda_0(y_{ij}) \exp(\eta_{ij}) \}$$

와 $V_i = \log(U_i)$ 에 대한 로그 가능도 함수인

$$\ell_{2i} = \log f(\alpha; v_i)$$

로 정의된다. 이 때, $v_i = \log(u_i)$ 이라고 할 때

$$\eta_{ij} = x_{ij}^T \beta + v_i$$

는 로그 위험률에 대한 선형 예측식 (linear predictor)이며, $\lambda_0(\cdot)$ 은 함수적인 형태가 알려져 있지 않으므로 회귀모수에 대한 추론을 위하여 기저 누적위험함수 $A_0(t)$ 를 다음과 같이 가정한다 (Ha 등, 2001):

$$A_0(t) = \int_0^t \lambda_0(k) dk = \sum_{k: y_{(k)} \geq t} \lambda_{0k}.$$

또한, 식 (2.2)의 다단계 가능도와 간단한 적분을 이용하여 다음과 같은 주변 가능도 (marginal likelihood)를 얻을 수 있다:

$$m = m(\beta, \lambda_0, \alpha) = \sum_i \log \left\{ \int \exp(h_i) d_{v_i} \right\},$$

여기서 $h_i = \sum_j \ell_{1ij} + \ell_{2i}$ 는 i 번째 개체의 다단계 가능도이며, 다음으로, 프레일티 $U_i = u_i$ 로 주어졌을 때 아래의 Breslow's 혹은 편 가능도 (partial likelihood)를 유도할 수 있다:

$$\ell_p = \sum_{ij} \delta_{ij} \eta_{ij} - \sum_k d_{(k)} \log \left\{ \sum_{(i,j) \in R_{(k)}} \exp(\eta_{ij}) \right\}, \quad (2.3)$$

여기서 $y_{(k)}$ 는 y_i 들 중에서 중복을 제외한 k 번째 ($k=1, \dots, D$) 순서의 사건 발생 시간, $d_{(k)}$ 는 시점 $y_{(k)}$ 에서의 사건이 발생한 수이며,

$R_{(k)} = R_{y_{(k)}} = \{(i, j) : y_{ij} \geq y_{(k)}\}$ 는 $y_{(k)}$ 에서의 위험에 노출된 집합 (risk set)을 의미한다.

더 나아가, 식 (2.3)의 편 가능도에 대해 벌점요소 (penalty term)를 추가하여 벌점화 편 가능도 (penalized partial likelihood (PPL); Ripatti와 Palmgran, 2000)를 다음과 같이 정의할 수 있다 (Kwon 등, 2018):

$$PPL = \ell_p - \frac{1}{2\alpha} \sum_i^q v_i^2$$

2.4 자료설명

본 연구에서는 2000년부터 2014년까지의 암 등록자료를 부산지역 암센터로부터 제공받아 분석에 활용하였다. 해당 자료는 연구종료 후에도 추가적인 추적조사를 통해 2016년까지의 환자 사망여부를 확인하여 추가한 것으로 나타났다. 자료에서 암 관련 지표 산출시 사용되는 국가 암 등록자료 주소가 읍면동 단위까지 상세하게 나타나있지 않거나, 병기를 알 수 없는 경우를 제외한 후 총 146,434명의 암 환자에 대하여 분석을 진행하였다. 전체 암 환자 중에서 주요 암에 대하여 위암 (C16) 환자는 24,025명, 대장암 (C18-C20) 환자 18,557명, 폐암 (C33-C34) 환자 13,967명 및 간암 (C22) 환자 14,063명 등으로 나타났다. 환자의 암 사망시간, 즉, 생존시간은 그 환자가 처음으로 암을 판정받은 날짜인 초진일부터 해당 암이 원인이 되어 사망한 날짜인 사망일까지로 정의하였으며, 이 때 분석단위는 해석의 용이성을 위해 평균 365.25일의 연도로 하였다. 연구종료시점인 2016년 12월 31일을 기준으로, 전체적인 중도절단 비율 (censoring rate)은 약 57.9%로서 절반보다 약간 많은 정도였다.

분석에 사용한 변수에 대한 분류는 다음과 같다. 기본적으로 암 환자의 인구학적 특성을 나타내는 성별과 연령 변수가 있으며, 생존시간을 산출하기 위한 초진일과 사망일, 그 외에 공변량으로 포함될 요약병기와 박탈지수 변수 등이 있다 (표 1). 분석에 사용한 변수에 대한 상세한 설명과 범주는 다음과 같다. 일반적으로 암 관련 지표를 산출할 때, 암 발생 또는 사망이 비교적 많지 않은 젊은 층의 연령 범주를 45세 미만인 0-44세로 고려하며 (Lee, 2008), 노인 연령에서 다시 65세를 기준으로 건강지표에 급격한 변화를 보인 연구가 많이 존재한다 (Hutchins, 1999). 또한, 본 등 록자료에서 고령 중에서도 연령이 75세 이상인 경우, 암 환자 대부분이 사망한 것으로 나타났으므로 연령군을 0-44세, 45-64세, 65-74세 및 75세 이상의 4개 범주로 분류하는 것이 합리적이라고 판단하였다. 요약병기 (summary stage)의 경우 병기 분류 방법인 미국 국립암연구소의 SEER (Surveillance Epidemiology and End Results) 병기 규칙은 암이 발생 지점으로부터 퍼진 정도에 따라 분류하는 가장 기본적인 방법 중 하나로 5개 정도의 주요 범주가 존재한다 (Young 등, 2001). 본 연구에서는 표와 그래프에 대한 보다 직관적이고 효율적인 해석을 위해 병기를 전이 정도가 가장 적은 순으로 국한 (localized), 국소 (regional) 또는 원격전이 (distant) 3개 범주로 분류하여 비교하였다.

소지역에 대한 분석의 편의를 위하여 15개년도의 부산광역시 읍면동을 임의로 105개로 통합하여 분석을 진행하였다. 또한, 환자의 국가 암 등록 자료 주소지에 따른 읍면동별 2010년 박탈지수 (deprivation index)를 고려하여 추가 분석을 시행하였다. 여기서 박탈지수는 통계청의 2010년 인구주택총조사의 10% 표본자료로부터 타지주택소유가구율, 1인 가구비율 등 11개의 지표를 활용하여 산출된 건강지표이다. 즉, 박탈지수는 주거 환경에 대한 삶의 질을 수치화하여 나타낸 것으로 각 지역의 사회·경제적

결핍수준을 반영하는 지표를 의미하며 (Townsend 등, 1988; Choi 등, 2011), 박탈지수 값이 큰 지역의 사회·경제적인 수준이 상대적으로 낮은 지역이라고 볼 수 있다.

표 1. 암 등록자료 분석 변수 설명

사용 변수	설명
생존시간	암 초진일부터 사망까지의 시간 (단위: 연도)
중도절단 여부	1: 사망, 0: 중도절단 (2016년 12월 31일 기준)
성별	1:남성, 2: 여성
연령군	1: 0-44세 2: 45-64세 3: 65-74세 4: 75세 이상
요약병기	1: 국한 2: 국소 3: 원격전이
동	2014년 기준 105개 읍면동으로 통합
박탈지수	지역별 사회 · 경제적인 수준

2.5 분석방법

모든 분석은 암 종별로 시행하였으며, 분석 방법 및 순서는 다음과 같다. 먼저 단변량 생존분석에서 성별에 따른 연령군 및 병기별 1년, 2년 및 5년 암 생존율을 산출하였으며, 각 해당하는 카플란 마이어 생존곡선 (survival curves)을 통해 전반적인 추이를 살펴보고 범주의 구분 또는 증감에 따라 생존율을 비교하였다. 다음으로, 다변량 생존분석에서는 성별, 연령군과 요약병기에 대한 변수를 공변량으로, 동을 변량효과로 고려하여 자료를 준모수적 로그-정규 프레일티 모형 (semi-parametric lognormal frailty model)에 적합하였다. 추정 결과에 대하여 다양한 그래프와 질병 지도를 작성한 후, 추가적으로 모형에 박탈지수 변수를 포함하여 적합하고 비교하였다. 본 연구에서는 정확한 통계분석을 위해 SAS 9.4 통계프로그

램을 사용하였으며, 특히 프레일티 모형 적합에 대해서는 벌점화 편 가능성도 추론법을 사용하였다 (SAS PHREG 프로시저 등). 또한, 일반적으로 통계 검정에서 유의성 판단의 기준으로 사용되는 유의수준 $\alpha = 0.05$ 를 사용하였다.



III 결과

3.1 생존율 추정

분석에 활용한 암 등록자료에서 부산광역시 전체 암 환자 146,434명을 암 종별로 살펴보았을 때, 위암은 24,025명으로 전체 중 약 16.4%로 가장 많았으며 대장암은 18,557명으로 12.7%, 폐암은 13,967명으로 9.5%, 간암의 경우 14,063명으로 9.6%정도를 차지하였다. 암 종별로 표 1의 범주 분류 기준에 따라 분석대상을 병기 및 연령별 암 환자 분포를 성별로 나누어 살펴보았을 때, 위암, 대장암 및 간암은 각 범주에서 성별은 여성보다 남성이, 연령은 45-64세 환자가 가장 많았으며, 병기 중 국한으로 판정받은 경우가 가장 많았다. 이와 달리 폐암의 경우 원격전이로 갈수록 해당하는 환자가 많았다 (표 2-표 5).

일반적으로 암 진단을 받은 후 5년 이상 생존하게 되면 완치 판정을 받는 것으로 알려져 있으므로 성별, 연령군, 병기에 따라 1년, 2년 및 5년 암 생존율을 산출하였다. 모든 암 중에서 생존율은 전체적으로 남성이 여성에 비해 낮은 편이었고, 연령이 증가할수록 감소하였으며, 병기의 경우 국한, 국소, 원격전이의 순으로 급격하게 감소하며, 원격전이는 눈에 띄게 매우 낮은 수치를 보여 예후가 매우 좋지 않은 것으로 나타났다. 특히, 위암과 대장암에서 국한은 5년 생존율이 젊은 연령에서 90% 이상으로 매우 높았으며 75세의 경우 절반 미만으로 감소하였다.

세부적으로는 위암에서 환자가 가장 많은 범주에 해당하는 45-64세의 국한 남성의 1년 생존율은 87.1%로 높은 편이었으며, 0-44세의 국한 여

성의 1년 생존율이 98.7%로 가장 높은 반면 75세 이상의 원격전이 남성의 5년 생존율은 1.5%로 가장 낮았다. 전체 중도절단 비율은 약 61.5%였으며, 비교적 적은 연령층에서 국한의 대부분이 90% 정도로 매우 높은 편에 속하였고, 75세 이상의 원격전이어서 비율은 약 1%에 불과하였다 (표 2). 대장암에서 환자가 가장 많은 45-64세의 국소 남성의 1년 생존율은 71.8%로 보통이었으며, 0-44세의 국한 여성의 1년 생존율이 96.6%로 가장 높은 반면 75세 이상의 원격전이 남성의 5년 생존율은 4.1%로 가장 낮았다. 전체 중도절단 비율은 약 62.9%였으며, 위암과 마찬가지로 적은 연령층에서 국한이 90% 정도로 매우 높은 편에 속하였고, 75세 이상의 원격전이어서 비율은 4~6% 정도였다 (표 3). 폐암의 경우 앞의 위암이나 대장암과는 약간 다른 경향을 보였다. 동일한 범주에서 여성보다 남성의 생존율이 대부분 10% 이상 적었으며, 75세 이상에서 여성의 생존율이 급격히 떨어지면서 그 차이는 감소하였다. 환자가 가장 많은 경우는 65-74세의 원격전이 남성이었으며, 해당 5년 생존율은 3.2%로 낮았다. 0-44세의 국한 여성의 1년 생존율이 81.6%로 가장 높은 반면 75세 이상의 원격전이 남성의 5년 생존율은 0.7%로 가장 낮았다. 전체 중도절단 비율은 약 18.6%였으며, 범주 중 90% 이상은 존재하지 않았다 (표 4). 간암에서 가장 많은 환자가 해당되는 45-64세의 국한 남성의 1년 생존율은 71.1%로 보통이었으며, 45-64세의 국한 여성의 1년 생존율이 79.3%로 가장 높은 반면 75세 이상의 원격전이 남성의 5년 생존율은 0.6%로 매우 낮았다. 전체 중도절단 비율은 약 21.6%였으며, 각 범주에서 모두 50% 미만으로 나타났다 (표 5).

카플란-마이어 생존곡선은 생존율 표와 전체적인 경향이 동일하여 다음과 같이 위암 결과만 나타냈으며 (그림 1), 그 외의 암 종에 대한 생존곡선은 부록 B에 있다 (그림 B.1-그림 B.3). 위암은 병기가 국한 또는 국

소일 때 0-44세에서 생존율이 높은 반면에 원격전이의 경우 연령에 상관 없이 생존율이 매우 낮았다. 곡선 오른쪽 상단에 나타난 로그순위 검정 (log-rank test) 시행 결과로부터 고려할 수 있는 모든 범주에서 병기 간의 생존율 차이가 매우 유의한 것으로 나타났다 ($p < .001$).



표 2. 성별에 따른 연령별 및 병기별 위암 생존율

	0-44세			45-64세			65-74세			75세 이상		
	국한 (N=600)	국소 (N=324)	원격전이 (N=158)	국한 (N=5,182)	국소 (N=2,356)	원격전이 (N=971)	국한 (N=2,864)	국소 (N=1,436)	원격전이 (N=530)	국한 (N=1,025)	국소 (N=519)	원격전이 (N=273)
중도절단 비율	90.8%	56.2%	5.7%	87.1%	53.4%	5.8%	72.8%	38.0%	3.2%	47.4%	17.7%	0.7%
1년	97.5%	88.6%	39.2%	96.9%	86.0%	37.9%	92.0%	74.3%	29.3%	79.3%	54.0%	16.5%
2년	95.3%	77.8%	21.5%	95.1%	73.1%	19.0%	88.1%	60.0%	13.8%	70.5%	36.6%	5.1%
5년	94.2%	61.5%	7.5%	90.6%	57.7%	7.5%	78.5%	44.0%	3.5%	48.7%	18.4%	1.5%
	0-44세			45-64세			65-74세			75세 이상		
	국한 (N=527)	국소 (N=344)	원격전이 (N=254)	국한 (N=2,067)	국소 (N=963)	원격전이 (N=350)	국한 (N=1,242)	국소 (N=525)	원격전이 (N=194)	국한 (N=709)	국소 (N=425)	원격전이 (N=187)
중도절단 비율	94.9%	65.4%	3.9%	93.0%	63.5%	5.4%	81.9%	47.8%	5.7%	45.3%	20.2%	1.1%
1년	98.7%	91.3%	44.1%	97.6%	90.3%	42.3%	94.4%	79.4%	24.2%	75.2%	53.2%	12.8%
2년	97.5%	82.0%	17.7%	96.2%	80.1%	17.7%	91.6%	65.5%	15.0%	65.4%	38.6%	5.4%
5년	95.4%	67.6%	5.3%	94.1%	66.5%	6.4%	85.6%	52.8%	7.2%	48.6%	22.4%	2.4%

표 3. 성별에 따른 연령별 및 병기별 대장암 생존율

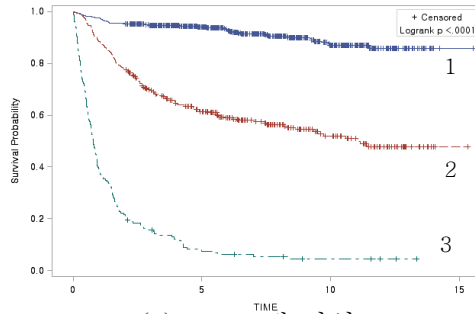
		0-44세			45-64세			65-74세			75세 이상		
		국한 (N=274)	국소 (N=295)	원격전이 (N=113)	국한 (N=2,219)	국소 (N=2,530)	원격전이 (N=897)	국한 (N=1,269)	국소 (N=1,631)	원격전이 (N=554)	국한 (N=499)	국소 (N=685)	원격전이 (N=262)
중도절단	비율	93.8%	76.6%	17.7%	88.9%	71.8%	18.0%	74.0%	62.2%	14.3%	50.7%	39.6%	3.8%
1년		98.2%	97.0%	66.4%	97.3%	95.7%	67.8%	93.7%	89.2%	54.9%	80.8%	80.7%	31.7%
2년		96.7%	89.8%	38.9%	95.9%	90.0%	42.5%	90.6%	82.5%	34.5%	71.9%	66.9%	16.4%
5년		94.7%	79.9%	18.2%	91.2%	75.4%	19.8%	79.9%	66.1%	14.5%	52.4%	42.4%	4.1%
		0-44세			45-64세			65-74세			75세 이상		
		국한 (N=206)	국소 (N=245)	원격전이 (N=101)	국한 (N=1,238)	국소 (N=1,455)	원격전이 (N=562)	국한 (N=713)	국소 (N=977)	원격전이 (N=360)	국한 (N=499)	국소 (N=709)	원격전이 (N=264)
중도절단	비율	96.6%	76.7%	26.7%	92.3%	78.2%	22.2%	80.8%	67.5%	13.9%	46.9%	42.7%	6.1%
1년		99.0%	96.7%	83.2%	98.4%	96.4%	74.0%	95.1%	93.5%	54.4%	78.6%	79.8%	31.1%
2년		98.5%	91.8%	60.4%	97.0%	91.1%	50.0%	92.3%	85.7%	29.2%	70.7%	64.0%	14.4%
5년		96.8%	80.5%	25.7%	93.6%	80.9%	22.0%	84.2%	72.7%	15.6%	50.8%	45.5%	6.0%

표 4. 성별에 따른 연령별 및 병기별 폐암 생존율

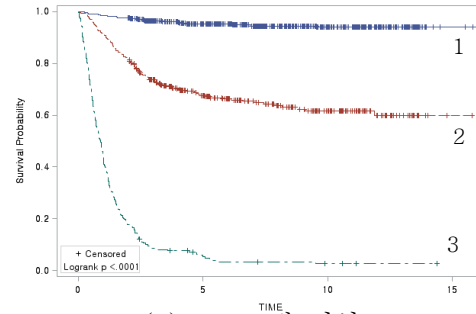
	0-44세			45-64세			65-74세			75세 이상		
	국한 (N=41)	국소 (N=49)	원격전이 (N=119)	국한 (N=793)	국소 (N=1,203)	원격전이 (N=1,853)	국한 (N=871)	국소 (N=1,302)	원격전이 (N=1,893)	국한 (N=416)	국소 (N=569)	원격전이 (N=958)
중도절단 비율	56.1%	53.1%	12.6%	47.5%	28.1%	4.6%	31.0%	18.1%	3.1%	16.6%	8.4%	0.8%
1년	78.1%	85.7%	54.6%	76.7%	71.2%	39.2%	66.0%	55.6%	24.7%	46.2%	36.0%	15.7%
2년	61.0%	71.4%	28.6%	64.3%	49.7%	17.3%	53.0%	35.0%	10.5%	29.1%	19.5%	5.0%
5년	55.1%	56.3%	9.1%	51.3%	31.2%	5.0%	33.3%	18.3%	3.2%	14.6%	8.2%	0.7%
	0-44세			45-64세			65-74세			75세 이상		
	국한 (N=49)	국소 (N=36)	원격전이 (N=110)	국한 (N=401)	국소 (N=349)	원격전이 (N=738)	국한 (N=315)	국소 (N=308)	원격전이 (N=589)	국한 (N=229)	국소 (N=244)	원격전이 (N=532)
중도절단 비율	81.6%	41.7%	10.9%	76.6%	47.6%	10.4%	53.7%	33.1%	9.7%	19.2%	16.4%	3.8%
1년	91.8%	77.8%	59.1%	94.8%	86.8%	59.5%	81.3%	69.5%	43.1%	43.7%	46.7%	21.4%
2년	89.8%	66.7%	37.3%	89.8%	71.9%	36.5%	70.8%	56.5%	26.5%	32.3%	28.7%	9.2%
5년	81.0%	43.5%	9.1%	79.3%	49.8%	8.7%	56.1%	33.1%	7.7%	19.8%	13.7%	3.8%

표 5. 성별에 따른 연령별 및 병기별 간암 생존율

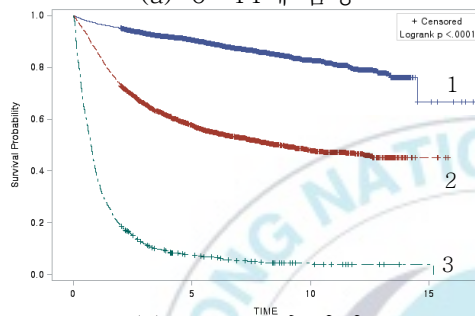
	0-44세			45-64세			65-74세			75세 이상		
	국한 (N=390)	국소 (N=213)	원격전이 (N=137)	국한 (N=3,582)	국소 (N=1,747)	원격전이 (N=1,018)	국한 (N=1,354)	국소 (N=682)	원격전이 (N=464)	국한 (N=544)	국소 (N=294)	원격전이 (N=203)
중도절단 비율	37.2%	12.7%	2.2%	34.7%	16.0%	2.7%	27.0%	11.3%	1.5%	12.9%	7.8%	0.5%
1년	65.4%	39.3%	16.8%	71.1%	47.1%	23.9%	65.9%	43.7%	20.0%	46.5%	31.6%	16.3%
2년	55.4%	25.8%	8.0%	59.5%	32.9%	11.7%	53.0%	28.7%	8.6%	29.6%	17.4%	4.4%
5년	43.1%	16.0%	3.7%	39.1%	17.9%	4.4%	29.4%	10.6%	1.2%	13.5%	6.3%	0.6%
	0-44세			45-64세			65-74세			75세 이상		
	국한 (N=84)	국소 (N=40)	원격전이 (N=24)	국한 (N=802)	국소 (N=341)	원격전이 (N=273)	국한 (N=627)	국소 (N=301)	원격전이 (N=169)	국한 (N=407)	국소 (N=209)	원격전이 (N=158)
중도절단 비율	40.5%	25.0%	8.3%	46.5%	16.7%	3.3%	25.8%	12.0%	4.1%	14.5%	4.8%	3.8%
1년	70.2%	57.5%	41.7%	79.3%	46.6%	26.0%	67.3%	44.5%	23.7%	46.7%	24.9%	9.5%
2년	57.1%	42.5%	25.0%	67.6%	32.3%	13.2%	52.2%	28.2%	13.0%	35.1%	12.4%	6.3%
5년	41.7%	21.6%	8.3%	51.9%	19.6%	3.5%	30.0%	11.4%	4.2%	14.5%	4.3%	3.5%



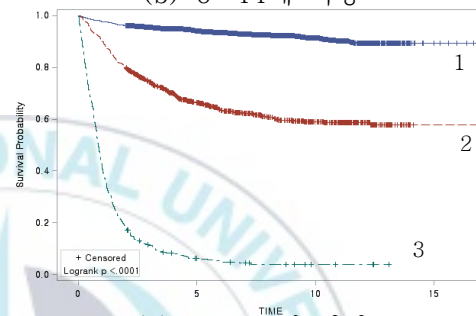
(a) 0-44세 남성



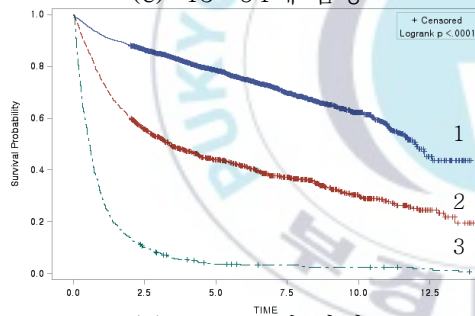
(b) 0-44세 여성



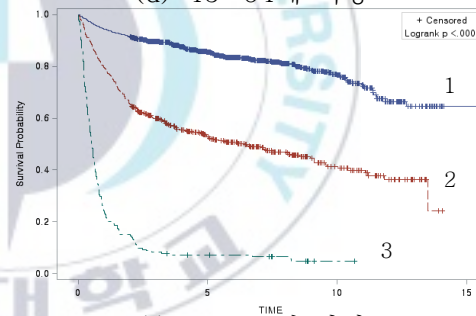
(c) 45-64세 남성



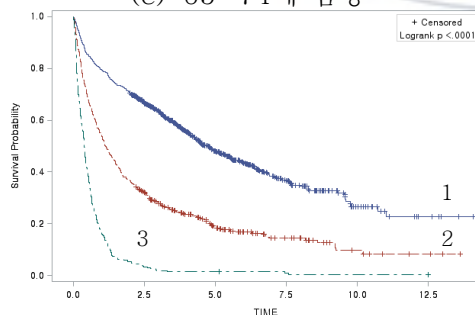
(d) 45-64세 여성



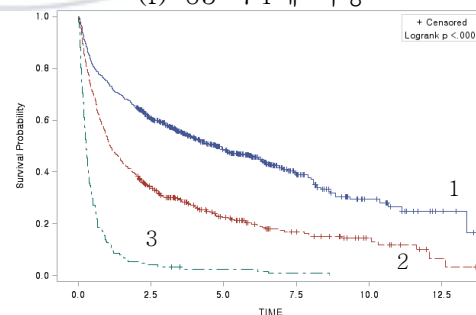
(e) 65-74세 남성



(f) 65-74세 여성



(g) 75세 이상 남성



(h) 75세 이상 여성

그림 1. 성별 및 연령에 따른 병기별 위암 카플란 마이어 생존곡선
(1: 국한, 2: 국소, 3: 원격전이)

3.2 모형적합 및 시각화 결과

다음으로 성별, 연령, 요약병기를 변량효과로 읍면동을 고려한 프레일티 모형에 적합하였다 (표 6-표 9). 그 결과 모든 암 중에서 여성보다는 남성일 때, 연령이 증가할수록, 전이 정도가 심할수록 암 사망에 대한 위험률이 전반적으로 유의하게 높은 것으로 나타났다. 보다 직관적인 해석을 위해 위험비 (hazard ratio; HR)를 암 종별로 살펴보았을 때, 다음과 같이 위암은 대장암과, 폐암은 간암과 비슷한 양상을 보였다. 먼저 위암에서 여성에 비해 남성이 약 $\exp(0.172) \approx 1.2$ 배, 0-44세에 비해 75세 이상인 경우 약 $\exp(1.436) \approx 4.2$ 배 위암으로 사망할 위험률이 더 높았다. 마찬가지로 요약병기의 국한에 비해 국소의 위암 위험률이 약 $\exp(1.227) \approx 3.4$ 배, 원격전이는 $\exp(2.801) \approx 16.5$ 배로 매우 높았다 (표 6). 대장암의 경우 여성에 비해 남성이 약 $\exp(0.138) \approx 1.2$ 배, 0-44세에 비해 75세 이상인 경우 약 $\exp(1.492) \approx 4.4$ 배, 국한에 비해 국소가 약 $\exp(0.604) \approx 1.8$ 배, 원격전이는 $\exp(2.306) \approx 10.0$ 배 대장암으로 사망할 위험률이 더 높았다 (표 7). 폐암의 경우 여성에 비해 남성이 약 $\exp(0.427) \approx 1.5$ 배, 0-44세에 비해 75세 이상인 경우 약 $\exp(1.133) \approx 3.1$ 배, 국한에 비해 국소가 약 $\exp(0.438) \approx 1.6$ 배, 원격전이는 $\exp(1.283) \approx 3.6$ 배 폐암으로 사망할 위험률이 더 높았다 (표 8). 간암의 경우 여성에 비해 남성이 약 $\exp(0.099) \approx 1.1$ 배, 0-44세에 비해 75세 이상인 경우 약 $\exp(0.624) \approx 1.9$ 배, 국한에 비해 국소가 약 $\exp(0.628) \approx 1.9$ 배, 원격전이는 $\exp(1.254) \approx 3.5$ 배 간암으로 사망할 위험률이 더 높았다 (표 9). 여기서 위험비에 대한 95% 신뢰구간 (confidence interval; CI)이 1을 포함하지 않으면 유의한 차이가 존재하는 것으로 볼 수 있다.

표 6. 위암: 프레일티 모형 적합 결과

변수	추정값	S.E.	p	HR	95% CI for HR				
고정효과									
성별	남성	0.172	0.023	<.001	1.19	1.136-1.243			
	여성	Ref							
연령군	0-44세	Ref	0.041	0.122	1.07	0.983-1.155			
	45-64세	0.064							
	65-74세	0.598					<.001	1.82	1.674-1.974
	75세 이상	1.436					<.001	4.20	3.862-4.577
병기	국한	Ref	0.025	<.001	3.41	3.246-3.583			
	국소	1.227							
	원격전이	2.801					<.001	16.46	15.572-17.404
변량효과									
동	0.0084	0.0028							

S.E.: 표준오차; Ref: 기준범주; HR (hazard ratio): 위험비; CI: 신뢰구간

표 7. 대장암: 프레일티 모형 적합 결과

변수	추정값	S.E.	p	HR	95% CI for HR	
고정효과						
성별	남성	0.138	0.025	<.001	1.15	1.093-1.205
	여성	Ref				
연령군	0-44세	Ref				
	45-64세	0.174	0.060	0.004	1.19	1.058-1.338
	65-74세	0.652	0.060	<.001	1.92	1.704-2.159
	75세 이상	1.492	0.061	<.001	4.44	3.942-5.011
병기	국한	Ref				
	국소	0.604	0.033	<.001	1.83	1.714-1.951
	원격전이	2.306	0.034	<.001	10.03	9.378-10.729
변량효과						
동	0.0121	0.0039				

표 8. 폐암: 프레일티 모형 적합 결과

변수		추정값	S.E.	p	HR	95% CI for HR
고정효과						
성별	남성	0.427	0.022	<.001	1.53	1.468–1.600
	여성	Ref				
연령군	0–44세	Ref				
	45–64세	0.242	0.063	<.001	1.27	1.127–1.441
	65–74세	0.611	0.063	<.001	1.84	1.629–2.083
	75세 이상	1.133	0.064	<.001	3.11	2.740–3.520
병기	국한	Ref				
	국소	0.438	0.030	<.001	1.55	1.462–1.642
	원격전이	1.283	0.027	<.001	3.61	3.420–3.806
변량효과						
동		0.0031	0.0016			

표 9. 간암: 프레일티 모형 적합 결과

변수		추정값	S.E.	p	HR	95% CI for HR
고정효과						
성별	남성	0.099	0.023	<.001	1.10	1.056–1.154
	여성	Ref				
연령군	0–44세	Ref				
	45–64세	–0.040	0.041	0.332	0.96	0.887–1.041
	65–74세	0.198	0.043	<.001	1.22	1.120–1.327
	75세 이상	0.624	0.046	<.001	1.87	1.704–2.044
병기	국한	Ref				
	국소	0.628	0.022	<.001	1.87	1.793–1.957
	원격전이	1.254	0.026	<.001	3.50	3.332–3.683
변량효과						
동		0.0069	0.0023			

프레일티 모형적합을 통해 추정된 각 동별 프레일티 추정값에 대하여 다음과 같은 보다 다양한 시각화를 제시하였다. 먼저, 읍면동 간 이질성 존재여부를 시각적으로 살펴보고자 그림 2-그림 5와 같이 105개의 동별 로그 프레일티 추정값과 각각에 대응하는 프레일티의 95% 신뢰구간에 대한 그래프를 작성하였다 (Ha 등, 2017). 그래프에서 동은 좌측부터 행정구역 순으로 나열하였으며 신뢰구간이 0을 포함하지 않는 경우 붉은 선으로 표기하였다. 105개 중 대부분 동의 신뢰구간이 0을 포함하였으나 몇몇 동에서 신뢰구간이 0을 포함하지 않아 다른 지역과 유의한 차이를 보였다. 따라서 동별 암 사망에 대하여 위험도의 차이인 이질성이 존재한다고 볼 수 있다.

위암에서 추정된 로그 프레일티 값이 매우 큰 남부민동, 반송동, 서동은 다른 지역보다 사망 위험이 높은 반면, 비교적 작은 추정값을 갖는 수민동과 화명동은 사망 위험이 낮았다 (그림 2). 대장암에서 다른 동에 비해 유일하게 사망의 위험이 높은 지역은 감천동, 낮은 지역은 좌동이었으며 (그림 3), 폐암에서 유일하게 유의한 차이를 보이는 좌동은 사망 위험이 낮은 지역이었다 (그림 4). 간암의 경우 사망의 위험이 높은 지역은 다대동, 낮은 지역은 우동, 좌동 및 구서동으로 나타났다 (그림 5).

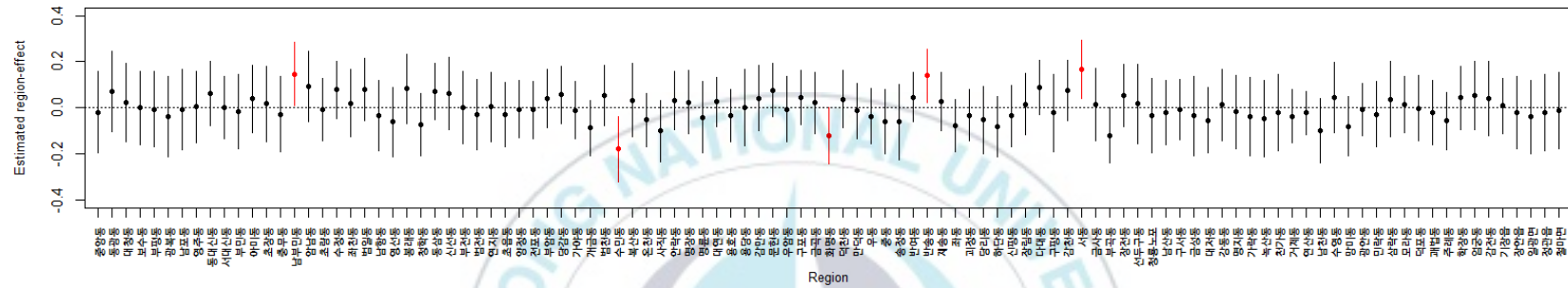


그림 2. 동별 위암의 로그 프레일티에 대한 95% 신뢰구간

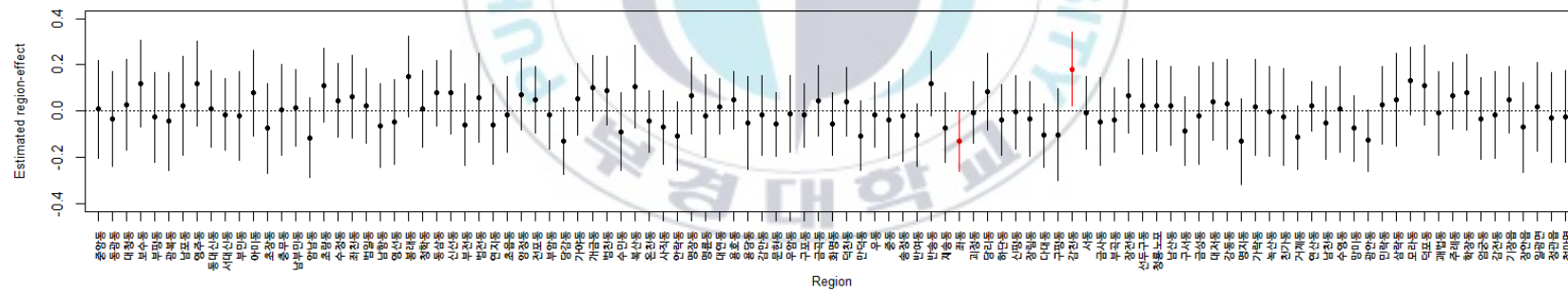


그림 3. 동별 대장암의 로그 프레일티에 대한 95% 신뢰구간

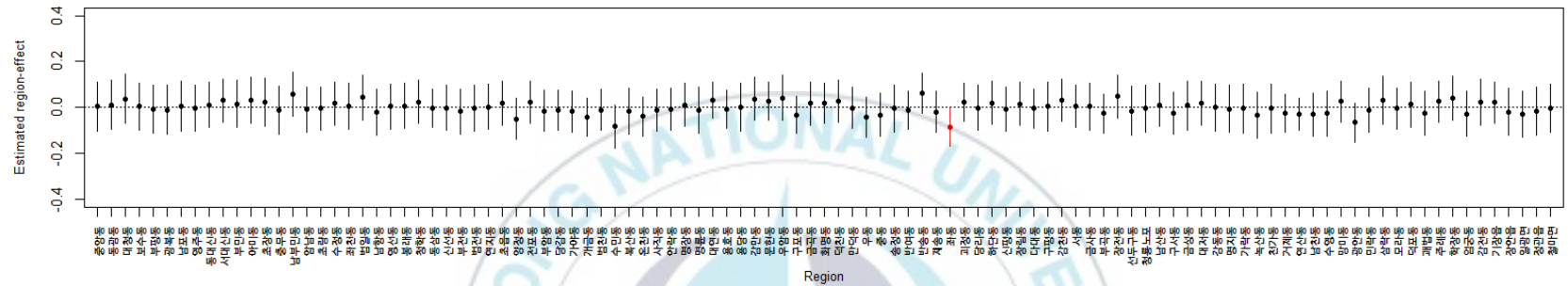


그림 4 동별 폐암의 로그 프레일티에 대한 95% 신뢰구간

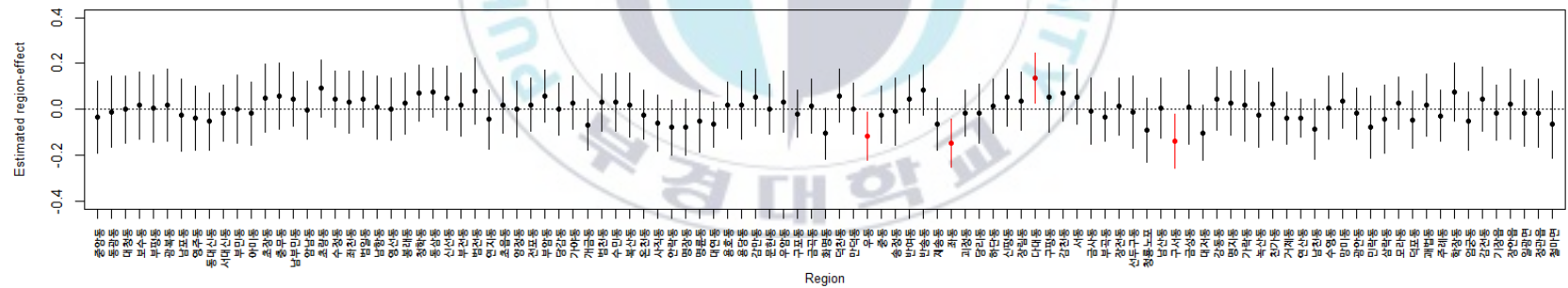


그림 5. 동별 간암의 로그 프레일티에 대한 95% 신뢰구간

암 종별로 첫 번째 모형적합 결과에서 가장 큰 프레이리티 추정값을 갖는 지역은 위암 서동, 대장암 감천동, 폐암 반송동 및 간암 다대동에 해당하였다. 이번에는 각각의 동에 대해 변량효과로 환자 개인을 고려하여 프레이리티 모형에 다시 적합하였다. 개인의 생존시간에 따라 모형적합을 통해 추정된 로그 프레이리티 값을 산점도로 표현하였다 (그림 6-그림 9). 이 때 중도절단여부를 서로 다른 표기를 사용하여 구분하였는데, 범례의 ‘Yes’ 가 사건이 발생 (즉, 사망)한 경우이며 ‘No’ 가 중도절단된 (즉, 생존) 경우를 나타낸다. 암 사망이 발생한 환자 대부분이 생존시간이 짧아 비교적 일찍 사건이 발생하였으며 로그 프레이리티 추정값이 전반적으로 0보다 큰 값으로 나타난 반면 생존한 암 환자의 프레이리티는 대체로 0보다 작은 값으로 추정되었으므로 상대적으로 적은 암 사망의 위험에 처한 것으로 볼 수 있다.

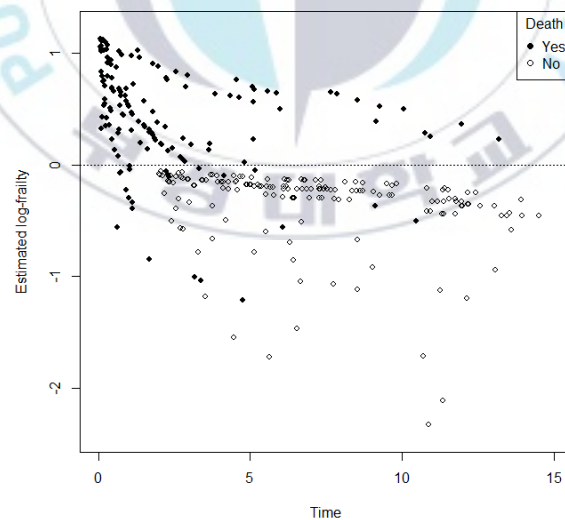


그림 6. 생존시간에 따른 위암 프레이리티 효과: 서동

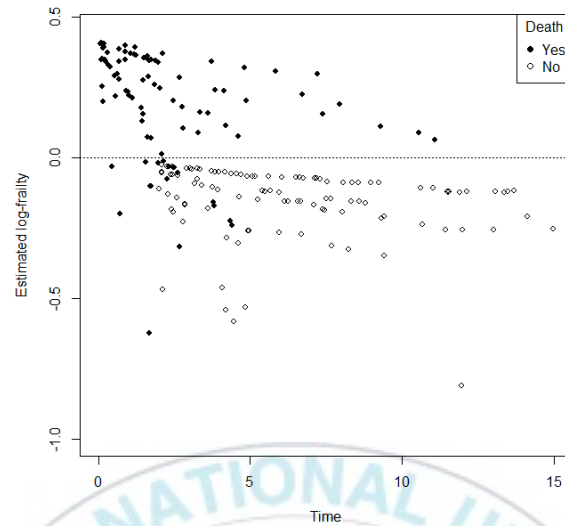


그림 7. 생존시간에 따른 대장암 프레일티 효과: 감천동

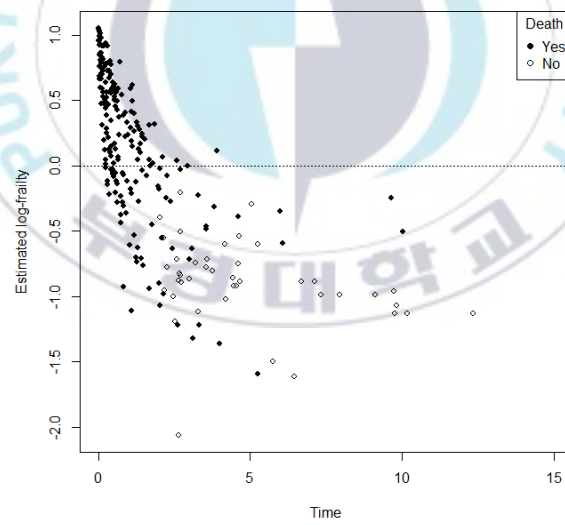


그림 8. 생존시간에 따른 폐암 프레일티 효과: 반송동

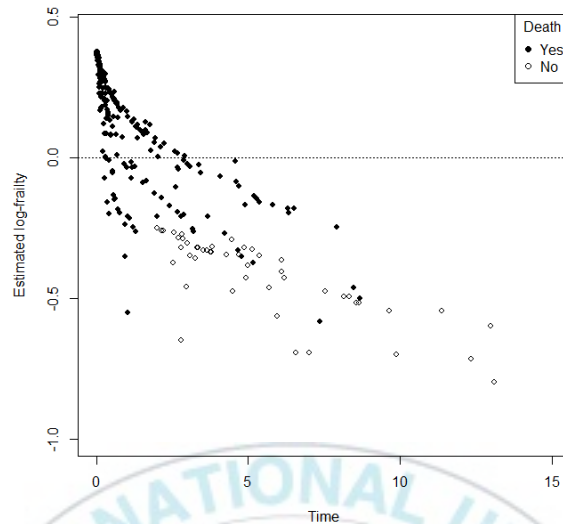


그림 9. 생존시간에 따른 간암 프레일티 효과: 다대동

그림 10-그림 13은 105개의 동별 로그 프레일티 추정값을 9분위로 나눈 후 부산광역시 동 단위 질병지도 (disease mapping)를 작성한 것이다. 각 색 범주에 동일한 개수의 동이 분류되도록 분위를 사용하였으며, 우측 하단에 색 범주에 따른 추정값의 범위를 표기하였다. 그림 2-그림 5의 프레일티 95% 신뢰구간 그래프에서 다른 지역과 유의하게 차이를 보인 동명을 해당 지도 위에 표기하였으며, 붉은색이 진할수록 프레일티 추정값이 높아 암 사망에 대한 위험이 높은 지역임을 의미한다. 즉, 유의하게 높은 위험도를 보였던 동의 경우 가장 진한 색으로, 낮은 위험도를 보인 동의 가장 연한 색으로 표현된다.

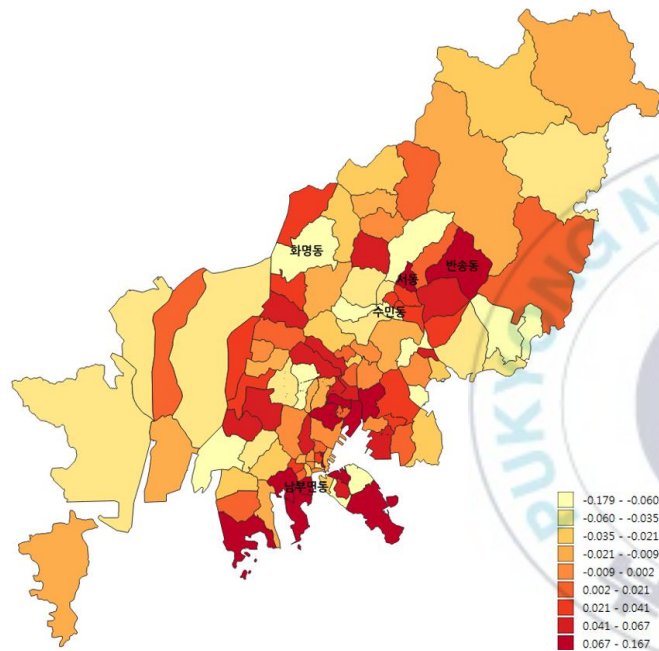


그림 10. 부산시 위암 질병지도 (9분위)

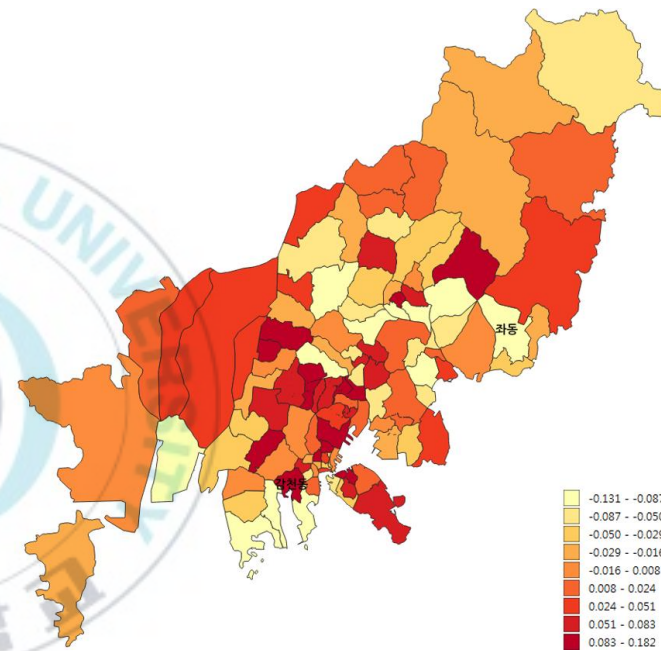


그림 11. 부산시 대장암 질병지도 (9분위)

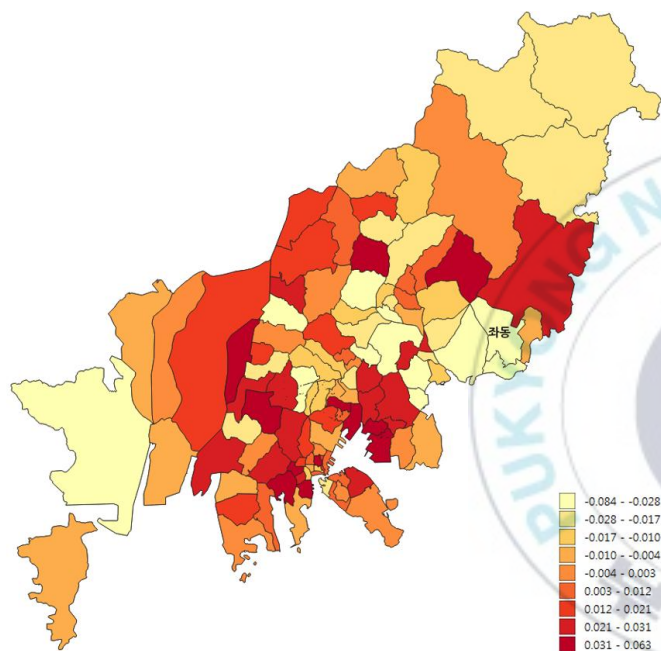


그림 12. 부산시 폐암 질병지도 (9분위)

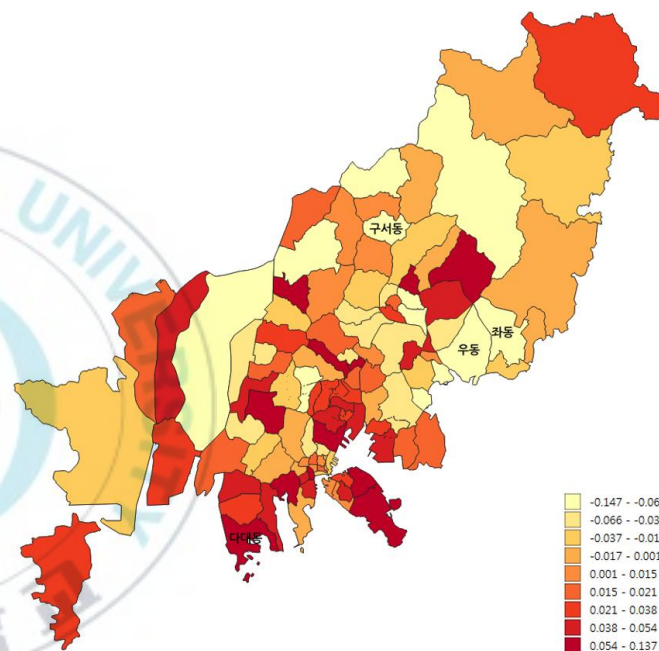


그림 13. 부산시 간암 질병지도 (9분위)

다음은 지역 박탈지수 값에 따른 로그 프레일티 추정값에 대한 암 종별 산점도를 나타낸 것으로, 단순회귀분석에서 박탈지수를 독립변수로, 프레일티 추정값을 종속변수로 두었을 때 추정된 회귀선을 추가하였다 (그림 14). 모든 암에서 추정된 회귀선이 양의 기울기를 보였으므로 박탈지수가 증가할수록 로그 프레일티 추정값이 선형으로 증가한다고 할 수 있다. 또한, 각 산점도의 왼쪽 상단에 표기되어있는 결정계수 R^2 을 통해 로그 프레일티 추정값에 대한 전체 변동 중에서 박탈지수는 위암에서 약 11.3%, 대장암은 14.5%, 폐암은 9.3%, 그리고 간암은 10.3% 정도를 설명하는 것으로 나타났다.

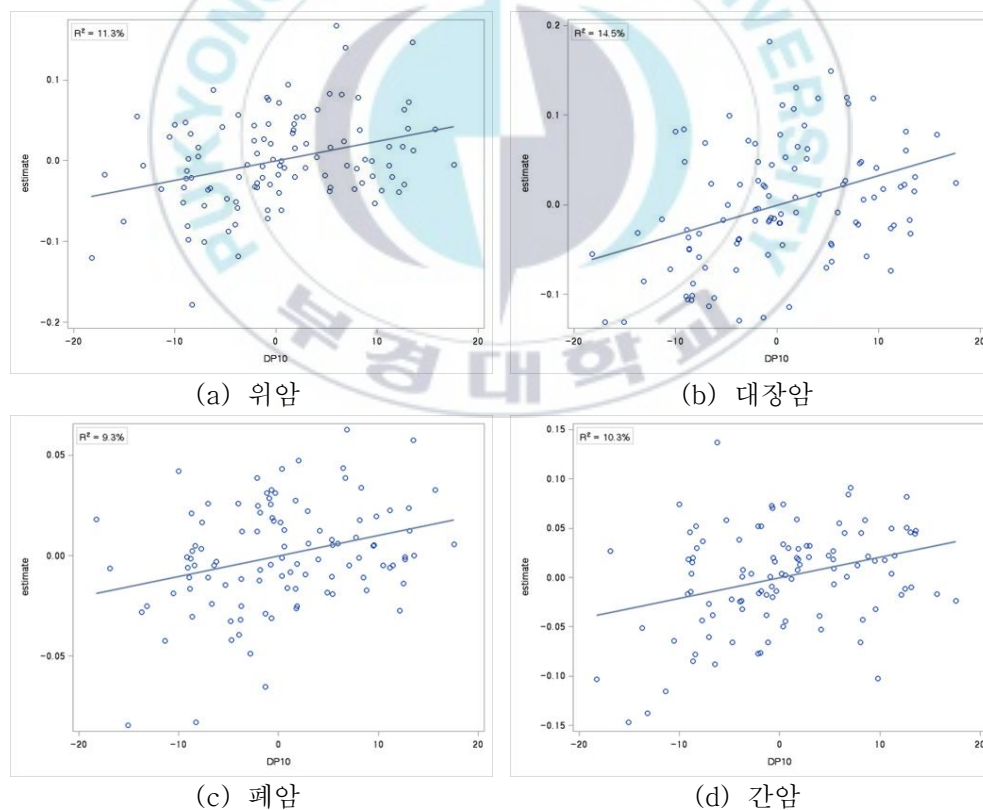


그림 14. 박탈지수에 따른 로그 프레일티 추정값

따라서 시각화 결과를 바탕으로 지역 박탈지수가 암 사망에 대한 위험도에 적지 않은 영향을 미칠 것이라고 판단하여 프레일티 모형에 연속형 변수인 박탈지수를 추가하여 적합하였다. 그 결과, 이전 모형과 비교하여 위험도 등의 추정값은 크게 변동이 없었다 (표 10-표 13).

표 10. 위암: 프레일티 모형 적합 결과: 박탈지수 포함

변수		추정값	S.E.	p	HR	95% CI for HR
고정효과						
성별	남성	0.172	0.023	<.001	1.19	1.135-1.242
	여성	Ref				
연령군	0-44세	Ref				
	45-64세	0.059	0.041	0.153	1.06	0.978-1.150
	65-74세	0.589	0.042	<.001	1.80	1.660-1.957
	75세 이상	1.427	0.043	<.001	4.17	3.826-4.534
병기	국한	Ref				
	국소	1.225	0.025	<.001	3.40	3.240-3.576
	원격전이	2.796	0.028	<.001	16.39	15.499-17.320
박탈지수		0.008	0.002	<.001	1.01	1.004-1.012
변량효과						
동		0.0047	0.0022			

S.E.: 표준오차; Ref: 기준범주; HR (hazard ratio): 위험비; CI: 신뢰구간

표 11. 대장암: 프레일티 모형 적합 결과: 박탈지수 포함

변수		추정값	S.E.	p	HR	95% CI for HR
고정효과						
성별	남성	0.140	0.025	<.001	1.15	1.095-1.207
	여성	Ref				
연령군	0-44세	Ref	0.060	0.004	1.19	1.055-1.334
	45-64세	0.171				
	65-74세	0.642				
	75세 이상	1.485				
병기	국한	Ref	0.033	<.001	1.83	1.711-1.947
	국소	0.602				
	원격전이	2.303				
박탈지수		0.011	0.002	<.001	1.01	1.006-1.015
변량효과						
동		0.0059	0.0029			

표 12. 폐암: 프레일티 모형 적합 결과: 박탈지수 포함

변수	추정값	S.E.	p	HR	95% CI for HR	
고정효과						
성별	남성	0.425	0.022	<.001	1.53	1.466-1.597
	여성	Ref				
연령군	0-44세	Ref	0.063	<.001	1.27	1.123-1.437
	45-64세	0.239				
	65-74세	0.605				
	75세 이상	1.126				
병기	국한	Ref	0.030	<.001	1.55	1.460-1.641
	국소	0.437				
	원격전이	1.283				
박탈지수		0.006	0.002	<.001	1.01	1.003-1.009
변량효과						
동		0.0015	0.0014			

표 13. 간암: 프레일티 모형 적합 결과: 박탈지수 포함

변수		추정값	S.E.	p	HR	95% CI for HR
고정효과						
성별	남성	0.098	0.023	<.001	1.10	1.055-1.153
	여성	Ref				
연령군	0-44세	Ref	0.308	0.96	0.885-1.039	
	45-64세	-0.042				
	65-74세	0.193				
	75세 이상	0.619				
병기	국한	Ref	<.001	1.87	1.791-1.956	
	국소	0.627				
	원격전이	1.252				
박탈지수		0.007	0.002	<.001	1.01	1.003-1.010
변량효과						
동		0.0043	0.0019			

박탈지수 변수를 포함하지 않은 경우와 마찬가지로, 다음과 같이 암 종에 따라 동별 프레일티의 95% 신뢰구간을 살펴보았다 (그림 15-그림 18). 위암, 대장암 및 폐암에서는 박탈지수를 포함하였을 때 동별 프레일티의 95% 신뢰구간이 모두 0을 포함하였으므로 모든 동이 유의한 차이를 보이지 않았다 (그림 15-그림 17). 간암에서는 다른 지역과의 사망 위험도가 유의하게 높은 다대동과 낮은 대저동이 존재하였으나 (그림 18), 박탈지수를 모형에 포함하지 않았을 때와 비교하였을 때 (그림 5) 유의한 차이를 보인 동의 개수가 절반으로 감소하였다. 따라서 박탈지수를 모형에 포함시키지 않았을 때보다 변량효과인 동의 유의성이 감소하였다고 할 수 있다.

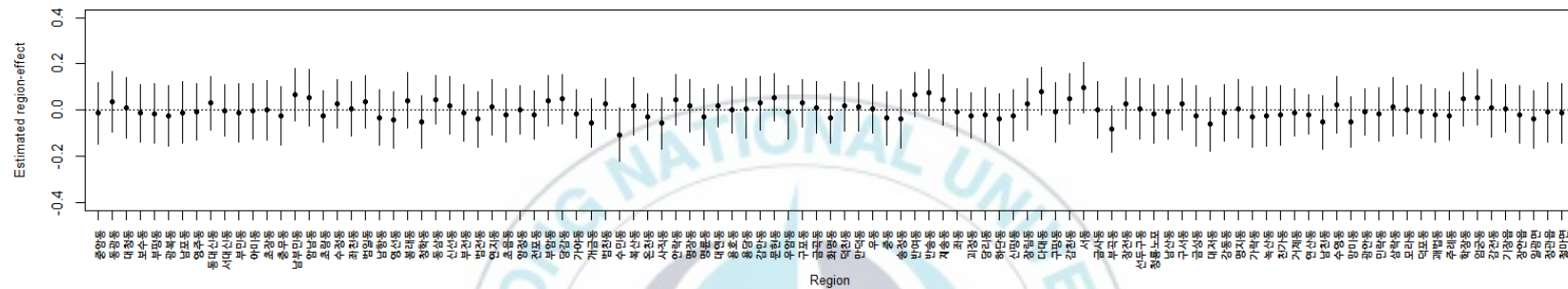


그림 15. 동별 위암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함

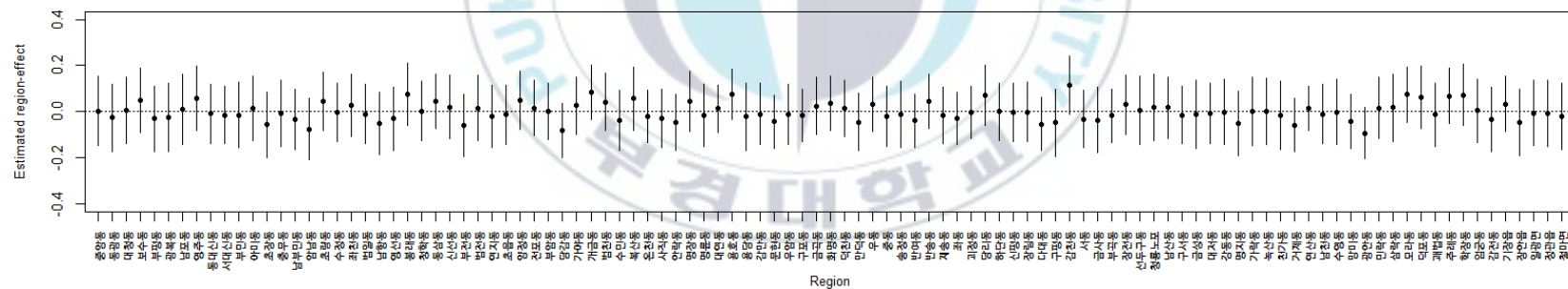


그림 16. 동별 대장암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함

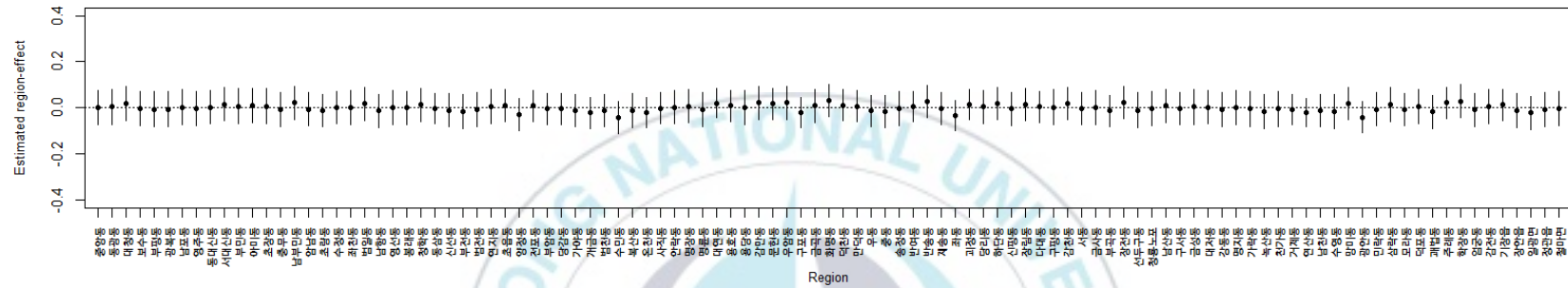


그림 17. 동별 폐암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함

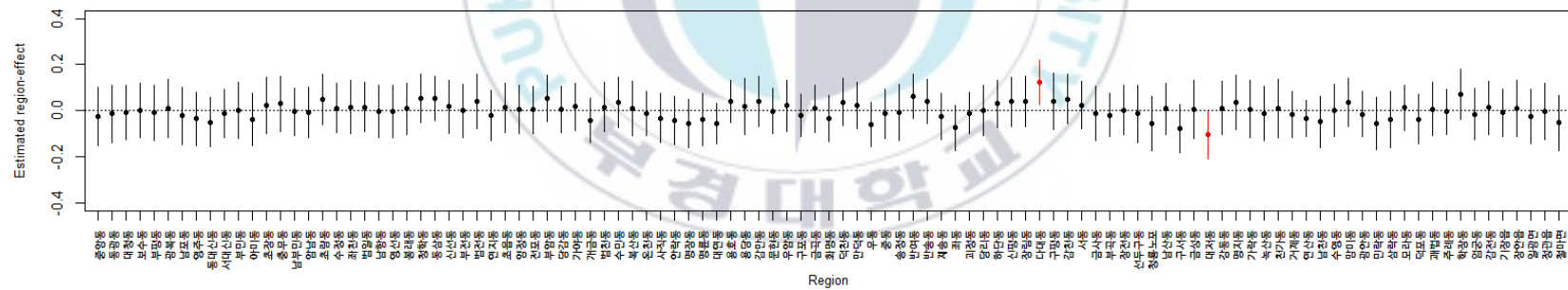


그림 18. 동별 간암의 로그 프레일티에 대한 95% 신뢰구간: 박탈지수 포함

IV 요약 및 결론

본 연구에서는 우리나라 주요 4대 암인 위암, 대장암, 폐암 및 간암에 대하여 암 생존자료를 토대로 분석을 시행하였다. 환자의 암 초진일부터 사망까지의 생존시간에 대한 생존율을 추정하기 위해 카플란-마이어 비모수적 방법을 이용하여 생존곡선을 작성하고 병기에 따른 생존율 차이에 대하여 로그 순위 검정을 시행하였다. 그 결과, 시간이 경과함에 따라 생존율이 감소하였으며 여성보다는 남성인 경우, 연령이 증가할수록, 또는 전이 정도 (병기)가 심할수록 암 생존율이 낮은 것으로 나타났으며, 암 종에 따라 서로 다른 양상을 보이기도 하였다.

동 단위를 변량효과로 고려하여 프레이리티 모형에 자료를 적합한 결과, 로그 프레이리티에 대한 95% 신뢰구간에서 암 종별로 몇몇 동이 0을 포함하지 않았으므로 암 사망의 위험에 대하여 동 간 유의한 이질성이 존재함을 확인할 수 있었다. 또한, 해당 추정값에 대한 전체적인 분포를 살펴보기 위하여 부산지역 동 단위 질병지도를 작성하였으며, 이를 통해 암 사망의 위험이 눈에 띄게 많거나 적은 지역을 파악하였다. 그 외에 3장의 결과에서 제시한 다양한 시각화 근거를 토대로 박탈지수 변수를 공변량으로 포함시켜 프레이리티 모형에 적합하였을 때, 소지역 간 격차가 감소하는 것을 확인하였다. 따라서 박탈지수 변수가 암 사망의 위험에 대한 동 간 변동을 어느 정도 설명한다고 할 수 있다.

마지막으로 향후에는 어떤 한 지역에서 암이 발병한 환자라고 하더라도 후에 다른 시도나 읍면동 등의 타 지역으로 이동하여 치료나 수술을 받은 경우 추적연계조사를 통해 보다 정밀한 연구가 가능할 것으로 보인다.

참고문헌

- [1] Choi, M. H., Cheong, K. S., Cho, B. M., Hwang, I. K., Kim, C. H., Kim, M.H., Hwang, S. S., Lim, J. H. and Yoon, T. H. (2011). Deprivation and mortality at the town level in Busan, Korea: An ecological study. *Journal of Preventive Medicine and Public Health*, **44**, 242–248.
- [2] Duchateau, L. and Janssen, P. (2008). *The frailty model*. Springer, New York.
- [3] Ha, I. D., Lee, Y. and Song, J. K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**, 233–243.
- [4] Ha, I. D., Jeong, J. H. and Lee, Y. (2017). *Statistical modelling of survival data with random effects: h-likelihood approach*. Springer, Singapore.
- [5] Hutchins, L. F., Unger, J. M. Crowley, J. J., Coltman, C. A. and Albain, K. S. (1999). Underrepresentation of patients 65 years of age or older in cancer–treatment trials. *The New England Journal of Medicine*, **341**, 2061–2067.
- [6] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.
- [7] Kim, D. W., Ryu, M. H., Kim, T. Y., Heo, D. S., Bang, Y. J., Park, J. G. and Kim, N. K. (2003). A multivariate analysis of prognostic factors in colorectal cancer. *Korean Journal of Medicine*, **64**, 268–274.
- [8] Kim, Y. S. and Tae, Y. S. (2011). The influencing factors on quality of life among breast cancer survivors. *Journal of Korean Oncology*

- Nursing*, **11**, 221–228.
- [9] Klein J. P. and Moeschberger, M. L. (2003). *Survival analysis: statistical methods for censored and truncated data*. Springer, New York.
 - [10] Kwon, J., Park, E. and Ha, I. D. (2018). Fitting semi-parametric frailty models using recent frailty R packages. *Journal of the Korean Data & Information Science Society*, **29**, 583–591.
 - [11] Lee, J., Yun, J. S., Jeong, J. J., Nam, K. H., Chung, W. Y., Sho, E. Y. and Par, C. S. (2008). Follicular thyroid carcinoma: Clinicopathologic features, prognostic factors, and treatment strategy. *Journal of Korean Surgical Society*, **74**, 34–41.
 - [12] Lee, T. and Lee, M. (2017). Analysis of stage III proximal colon cancer using the Cox proportional hazards model. *Journal of the Korean Data & Information Science Society*, **28**, 349–359.
 - [13] Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society B*, **58**, 619–678.
 - [14] Ministry of Health and Welfare (2015). Annual report of cancer statistics in Korea in 2015.
 - [15] Paik, H. C., Maeng, D. H., Song, S. W., Kim, K. D. and Chung, K. Y. (2002). Surgery for lung metastases from colorectal cancer. *Journal of the Korean Society of Coloproctology*, **18**, 37–41.
 - [16] Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**, 1016–1022.
 - [17] Statistics Korea (2016). Cause of Death Statistics in 2017. Korean statistical information service (KOSIS). <http://kosis.kr/>
 - [18] Townsend, P., Phillimore, P. and Beattie, A. (1988). *Health and deprivation: inequality and the North*. Croom Helm, London.

- [19] Yoon, T. H. (2010). Regional health inequalities in Korea The Status and Policy Tasks. *Journal of Critical Social Welfare*, **30**, 49–77.
- [20] Young, J. L. Jr., Roffers, SD., Ries, LAG., Fritz, AG. and Hurlbut, AA. (2001). *SEER Summary staging manual: 2000 codes and coding instructions*. Bethesda, MD: National Cancer Institute.



부록 A.1

SAS 코드

```
LIBNAME L3 "(경로지정)";

%MACRO CLEAN1;
/* 병기: 알수없음 제거 */
%DO I=1 %TO 4;
DATA CANCER6;
SET L3.CANCER&I._2;
IF S_CODE=4 THEN DELETE;
RUN;

/* 박탈지수 MERGE */
PROC SORT DATA=CANCER6; BY DONG; RUN;
PROC SORT DATA=L3.C2; BY DONG; RUN;

DATA CANCER6_2;
MERGE CANCER6 (IN=IN1) L3.C2 (IN=IN2);
BY DONG;
IF IN1;
RUN;

/** 생존시간 **/
/* 사망일이 초진일과 맞지 않는 경우 -> 사망일 결측 */
DATA TEMP1;
SET CANCER6_2;
IF DEATH='11111111' THEN DEATH='';
RUN;

DATA TEMP1;
SET TEMP1;
IF DEATH='' THEN DEATH_1='20161231';
ELSE DEATH_1=DEATH;
RUN;

DATA SURV6_&I;
SET TEMP3;
YEAR=SUBSTR(FDX, 1, 4);
```

```

END=INPUT(DEATH_1, yymmdd8.);
TIME= (END - START) / 365.25;
IF DEATH=' ' THEN STATUS = 0; /* 중도절단 여부 */
ELSE STATUS = 1;
RUN;
%END;
%MEND;

/* 병기별 생존율 및 카플란마이어*/
%MACRO KM(DATA);
%DO I=1 %TO 2;
%DO J=1 %TO 4;
DATA KM&I&J;
SET &DATA;
IF SEX="&I" AND AGE_G="&J";
RUN;

PROC LIFETEST DATA = KM&I&J PLOTS = SURVIVAL(TEST) INTERVALS = 0
TO 5 BY 0.0001;
TIME TIME*STATUS(0);
STRATA S_CODE;
RUN;
%END;
%END;
%MEND ;
%KM(SURV6_1_2);

/** 다수준 생존분석 **/
/* frailty(random effect)는 동 */
%MACRO FRAILITY;
%DO I=1 %TO 4;
PROC PHREG DATA=SURV6_1_2;
    CLASS AGE_G(ref="1") SEX(REF='2') S_CODE(REF='1 국한')
DONG;
    MODEL TIME*STATUS(0)=AGE_G SEX S_CODE/RL;
    RANDOM DONG/SOLUTION ;
RUN;

PROC PHREG DATA=SURV6_1_2; /* 박탈지수 변수 추가 */
    CLASS AGE_G(ref="1") SEX(REF='2') S_CODE(REF='1 국한')
DONG;

```

```

MODEL TIME*STATUS(0)=AGE_G SEX S_CODE DP10/RL;
RANDOM DONG/SOLUTION;

RUN;
%END;
%MEND;

/* 암종별 프레일티가 가장 높은 동에 대하여
RANDOM EFFECT를 개인으로 두고 산점도 작성 (중도절단 구분) */
DATA SURV6_1_3;SET SURV6_1_2;IF DONG IN ("서동");RUN; *위암HIGH:
서동;
DATA SURV6_2_3;SET SURV6_2_2;IF DONG IN ("감천동");RUN; * 대 장 암
HIGH: 감천동;
DATA SURV6_3_3;SET SURV6_3_2;IF DONG IN ("반송동");RUN; *폐암HIGH:
반송동;
DATA SURV6_4_3;SET SURV6_4_2;IF DONG IN ("다대동");RUN; *간암HIGH:
다대동;

%MACRO INDV1;
%DO I=1 %TO 4;
ODS OUTPUT SOLUTIONR = RES&I;
PROC PHREG DATA=SURV6_&I._3;
CLASS SEX(REF="2") AGE_G(REF="1") S_CODE(REF="1 국한") SERIAL;
MODEL TIME*STATUS(0)=SEX AGE_G S_CODE/MAXITER=40;
RANDOM SERIAL/SOLUTION;
RUN;
ODS OUTPUT CLOSE;

/* 중도절단 여부를 알기 위해 위의 프레일티 결과와 MERGE */
PROC SORT DATA=RES&I;BY SERIAL;RUN;
PROC SORT DATA=SURV6_&I._3;BY SERIAL;RUN;

DATA INDV&I;
MERGE RES&I (IN=IN1) SURV6_&I._3 (IN=IN2 KEEP=SERIAL STATUS DONG
TIME);
BY SERIAL;
IF IN1;
RUN;
%END;
%MEND;
%INDV1;

```

부록 A.2

R 코드

1. 프레일티 95% 신뢰구간 그래프 작성

```
> setwd('(경로지정)')

> library(readxl)
> windowsFonts(malgun = windowsFont("나눔바른고딕"))

> data1 = read_excel('FRAILITY_TYPE_3.xlsx', sheet = 'GAST_1') # 데이터 불러
오기
> plot(data1$estimate[data1$lower <= 0 & data1$upper >= 0] ~ data1$code[
data1$lower <= 0 & data1$upper >= 0], ylim=c(-0.4,0.4), ylab="Estimated
region-effect", xlab = "Region", pch = 20, xaxt = 'n', xlim = c(4, 102),
cex.axis = 0.9, cex.lab = 0.8, cex = 0.95)
> points(data1$estimate[data1$lower > 0 | data1$upper < 0] ~ data1$code[
data1$lower > 0 | data1$upper < 0], pch = 20, cex = 0.95, col = 2) # 색
구분
> axis(1, at = data1$code, labels = data1$dong_2, las = 2, cex.axis = 0.7, family
= "malgun", font = 2)
> abline(h=0, lty=3)
> for (i in data1$code[data1$lower <= 0 & data1$upper >= 0]) {
  x1<- c(i, i)
  y1<- c(data1$lower[i], data1$upper[i])
  lines (y1 ~ x1)
}
> for (i in data1$code[data1$lower > 0 | data1$upper < 0]) {
  x1<- c(i, i)
  y1<- c(data1$lower[i], data1$upper[i])
  lines (y1 ~ x1, col = 2)
}
> data1$dong_2[(data1$lower>0 | data1$upper<0)] # 유의한 동 추출
```

2. 암 종별 가장 큰 프레이리티 추정값을 갖는 동별 산점도

```
> indv = read_excel('./FRAILITY_INDV_1.xlsx', sheet = 1) # 위암
> a = min(indv$Estimate); b = max(indv$Estimate)

> plot(indv$TIME, indv$Estimate, cex = 0.8, col = 1, pch = 16, xlab = 'Time',
      ylab = 'Estimated log-frailty')
> plot(indv$TIME[indv$STATUS == 1], indv$Estimate[indv$STATUS == 1], cex
      = 0.8, col = 1, pch = 16, ylim = c(a-.1, b+.1), xlim = c(0, 15), xlab =
      'Time', ylab = 'Estimated log-frailty')
> points(indv$TIME[indv$STATUS == 0], indv$Estimate[indv$STATUS == 0],
      cex = 0.8, col = 1, pch = 1)
> abline(h = 0, lty = 3)
> legend("topright", legend = c('Yes', 'No'), title = 'Death', pch=c(16, 1))

# 대장암, 폐암, 간암 코드 동일
```

부록 B

추가 그림

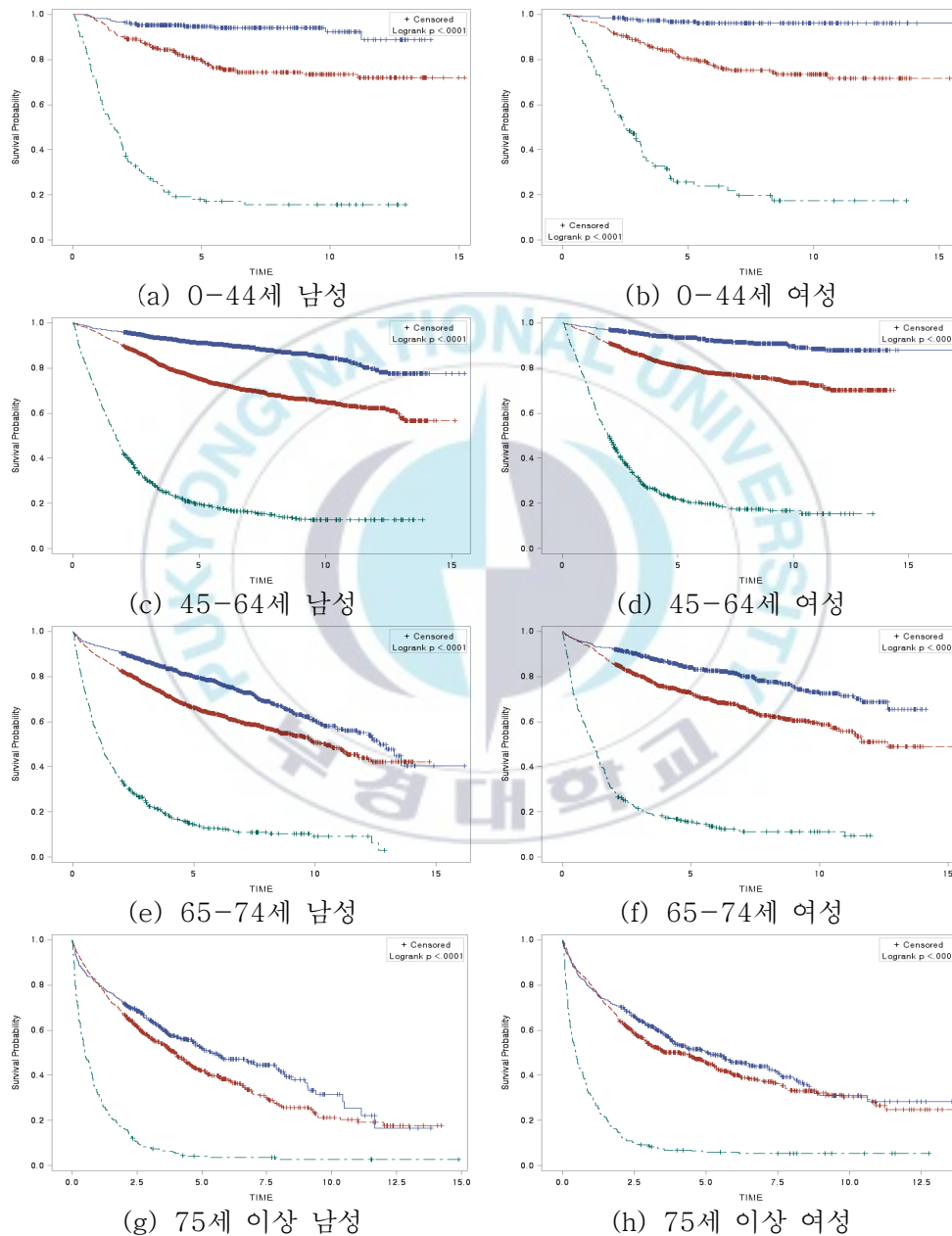
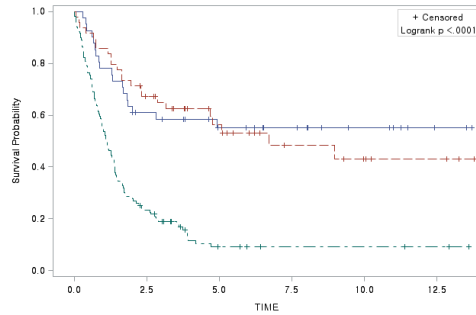
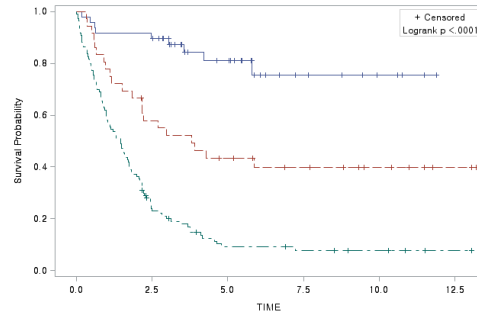


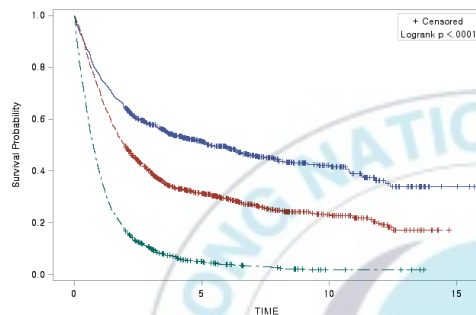
그림 B.1. 성별 및 연령에 따른 병기별 대장암 카플란 마이어 생존곡선



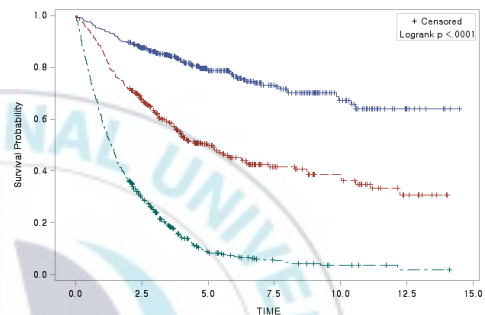
(a) 0-44세 남성



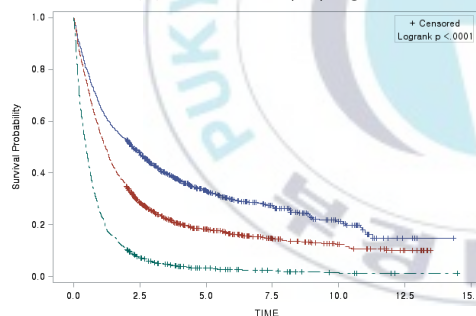
(b) 0-44세 여성



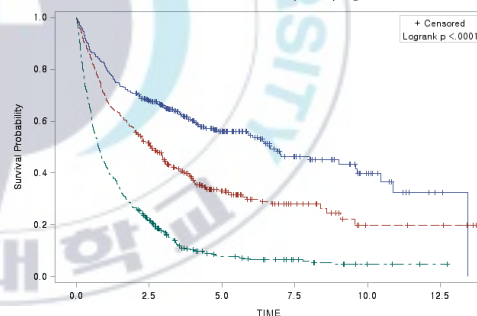
(c) 45-64세 남성



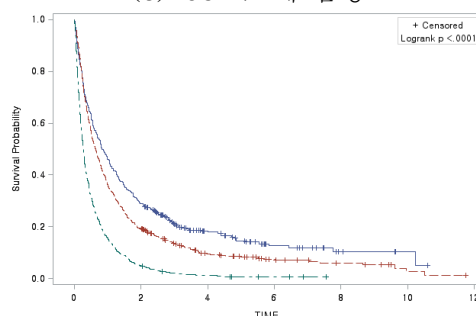
(d) 45-64세 여성



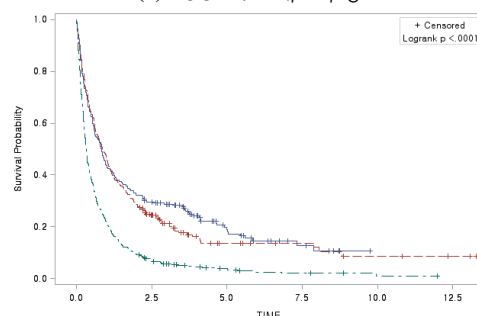
(e) 65-74세 남성



(f) 65-74세 여성

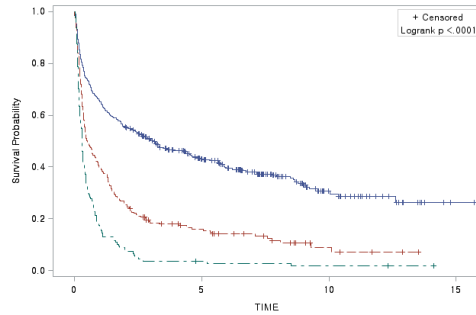


(g) 75세 이상 남성

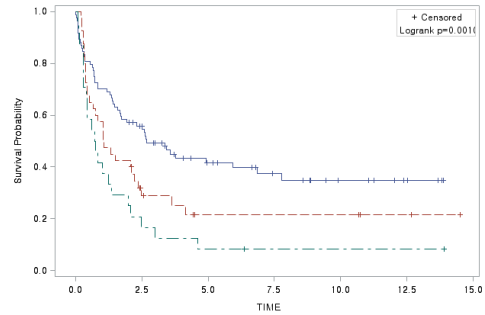


(h) 75세 이상 여성

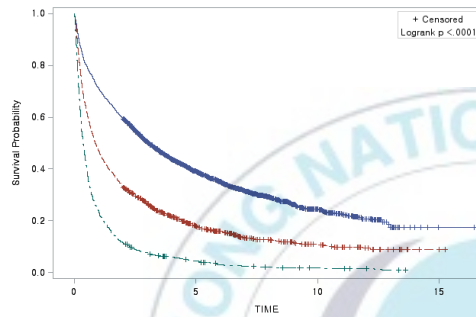
그림 B.2 성별 및 연령에 따른 병기별 폐암 카플란 마이어 생존곡선



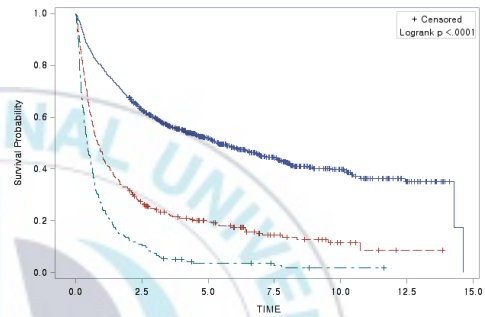
(a) 0-44세 남성



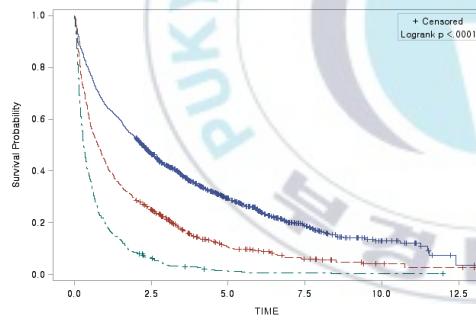
(b) 0-44세 여성



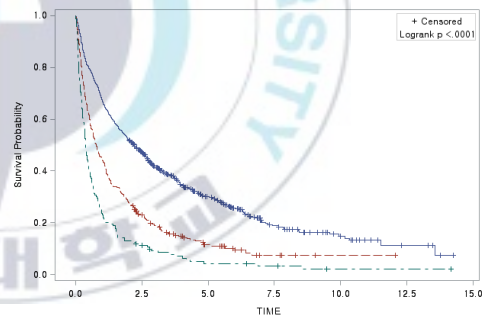
(c) 45-64세 남성



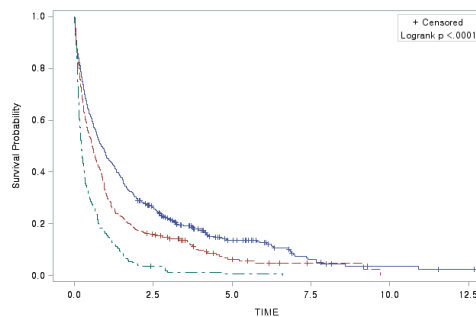
(d) 45-64세 여성



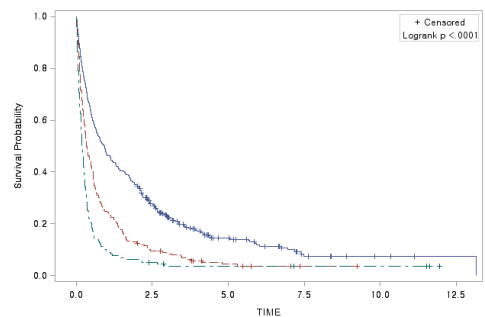
(e) 65-74세 남성



(f) 65-74세 여성



(g) 75세 이상 남성



(h) 75세 이상 여성

그림 B.3. 성별 및 연령에 따른 병기별 간암 카플란 마이어 생존곡선