왼쪽 절단된 코호트에서
시간 척도 선택에 따른 차이 비교 연구

연세대학교 보건대학원

보건통계학과 보건통계전공

단 비

왼쪽 절단된 코호트에서 시간 척도 선택에 따른 차이 비교 연구

지도 남 정 모 교수

이 논문을 보건학 석사학위 논문으로 제출함

2019년 12월

연세대학교 보건대학원
보건통계학과 보건통계전공
단 비

단비의 보건학 석사학위 논문을 인준함

심사위원 남 정 모 인

심사위원 박 소 희 인

심사위원 한 민 경 인

연세대학교 보건대학원

2019년 12월 일

감사의 글

수업 끝나고 집에 도착하면 새벽 1시가 다 되어가는 시간. 오송과 대전에서 서울을 오가며 몸은 고단했지만, 마음은 참 행복했습니다. 이제 곧 다가오는 졸업이 시원섭섭하지만 대학원 생활하면서 배운 것들이 저에게는 또 다른 도전을 시작하게 할 자양분이 되어 줄 것 같습니다.

뒤돌아보면 많은 분의 도움이 있었기에 대학원 생활을 잘 마무리할 수 있었습니다. 무엇보다도 먼 길을 오가며 학업의 열정을 지속할 수 있었던 이유는 애정을 가지고 지도해 주셨던 교수님들의 가르침 덕분이었습니다. 논문의 틀을 잡아주시고 연구자로서 거시적 시각을 갖도록 지도해주셨던 남정모 교수님, 통계학에 대한 통찰력은 저에게 학문적인 고민과 균형적인 생각을 할 수 있게끔 이끌어주셨습니다. 통계 수업을 이해하기 쉽게 알려주실 뿐만 아니라, 비판적인 시각의 눈을 기를 수 있도록 지도해주신 박소희 교수님, 수업마다 보여주셨던 열정 잊지 못할 것 같습니다. 마지막까지 세심하게 주말도 상관없이 지도해주셨던 한민경 박사님, 진심으로 감사의 말씀을 전하고 싶습니다. 모르는 것이 있으면 귀찮게 물어보았던 장현수 선생님을 비롯한 통계상담실 선생님들 도움 덕분에 대학원 과정 잘 마칠 수 있었습니다.

우리 최강통계 동기들 아람, 세은, 다경, 주리 언니에게도 고마운 마음을 전합니다. 함께하는 동기들이 있었기에 학교생활이 즐거웠고, 마지막까지 최선을 다할 수 있었습니다.

일과 학업을 병행함에 있어 많은 배려를 해주신 이진수 팀장님, 김수연 팀장님, 윤태영 팀장님, 장무영 선생님, 정태영 선생님, 윤은혜 선생님, 김서윤 선생님, 우한나 선생님, 백주은 선생님, 이루리 선생님께도 감사의 마음을 전합니다. 학교 진학을 고민할 때, 진심어린 조언을 해주셨던 박순만 단장님께도 깊은 감사의 말씀을 전합니다.

아울러, 그 누구보다도 딸의 선택을 진심으로 응원해주시는 아빠와 엄마, 그리고 항상 격려의 말을 아끼지 않으시는 시아빠와 시엄마, 제가 너무나 사랑하고 마음 깊이 존경합니다. 언제는 친구같이 때로는 오빠같이 조언해주는 동생

재혁에게도 고마운 마음을 표합니다. 만날 때마다 유쾌한 아주버님과 형님, 보고만 있어도 행복해지는 조카 하솜이 사랑하고, 감사합니다.

마지막으로 나의 가장 소중한 사람이자 사랑하는 남편, 지친 몸과 마음을 재충전할 수 있도록 늘 응원해주고 아낌없이 지지해주어 고맙습니다. 졸린 눈 비벼가며 픽업 와주던 사랑 잊지 않을게요.

2019년 12월
단비 올림

차 례

표 차 례

그 림 차 례

국문요약

왼쪽 절단된 코호트에서
시간 척도 선택에 따른 차이 비교 연구

연구 배경

시간 척도 선택은 역학 코호트 연구에서 생존 분석 시 중요한 문제이다. 콕스 비례위험모형(Cox's Proportional Hazard Model)을 사용하는 생존 및 관찰 연구에서는 연구에 등록되어 사건이 발생할 때까지의 시간을 계산한 연구 기간(Time-on study)이라는 시간 척도를 주로 사용해왔다. 연령은 보건학에서 질병 발생 위험에 영향을 주는 대표적인 혼란 변수로 일반적으로 모형에 보정하여 분석하고 있다. 그러나, 연구 기간으로 사용할 경우 연구에 같은 시점에 등록된 대상자들이 동일한 관측 기간이 주어진다 하여 발생 위험이 모두 동일하다고 할 수 있는지에 대한 의문이 제기된다. 또한, 국외 선행연구자들은 연구 대상자들이 각기 다른 시점과 연령에서 연구에 등록되어 왼쪽 절단된 형태가 발생하기 때문에 이는 생존 추정치를 증가시키고, 공변량의 추정된 효과를 편향시킬 수 있는 가능성을 제기하고 있다. 이에 태어난 0세부터 사건이 발생할 때까지의 시간을 계산한 도달 연령(Attained age) 시간 척도를 적용해봄으로써 시간 척도를 달리 선택함에 발생하는 편차(Bias)가 있을지 비교하고자 본 연구를 진행하였다.

대상 및 방법

유방암 발생 코호트와 위암 발생 코호트를 이용하여 시간 척도 모형을 비교

하였다. 유방암 발생 코호트는 폐경 전·후 여성에서 위험인자가 유방암 발생에 미치는 영향에 관한 주제였고, 위암 발생 코호트는 알코올성 간질환 유무에 따라 위암 발생에 미치는 영향에 관한 것이었다. 두 연구주제 모두 2002년부터 2013년까지 국민건강보험공단 검진코호트 DB를 이용하여 40세 이상 성인을 대상으로 하였다. 기저 위험 함수(Baseline hazard function)가 지수분포(Exponential distribution)를 따르는지 여부와 코호트 진입 연령과 공변량 간 독립 여부를 파악하였다. 콕스 비례위험모형을 이용하여 두 가지 여부에 따른 위험비(Hazard Ratio)와 95% 신뢰구간(95% Confidence Interval), 그리고 추정된 베타 계수 값의 변화 정도를 나타내는 Fraction 값을 산출하여 차이를 비교하였다. 분석에 사용된 모델은 총 4가지로 모델 1은 시간 척도를 연구 기간으로 한 모형, 모델 2는 시간 척도를 연구 기간으로 하고 연령을 보정한 모형, 모델 3은 시간 척도를 도달 연령으로 하고 출생연도를 5년 구간으로 층화한 모형, 마지막으로 모델 4는 시간 척도를 도달 연령으로 하고 왼쪽 절단 분석을 적용한 모형으로 구성하였다.

연구 결과

첫째, 왼쪽 절단된 코호트에서 연구 기간을 시간 척도로 사용한 것과 도달 연령을 시간 척도로 사용하고 왼쪽 절단 분석한 모델들 간 위험비 방향이 바뀌거나, 95% 신뢰구간 및 Fraction 값이 달라지고 있음을 확인하였다.

둘째, 기저 위험 함수가 지수분포를 따르거나 또는 코호트 진입 연령과 공변량 간 독립이면 두 시간 척도 분석 결과의 편차는 줄어드는 결과를 보였다. 그러나, 코호트 진입 연령과 공변량 간 독립일지라도 간혹 차이를 보이는 변수들이 유방암과 위암 발생 코호트에서 모두 존재했다.

셋째, 유방암 발생 코호트의 폐경 변수에서는 모델 2를 제외하고 모델 1, 3, 4에서는 다른 변수들(의료보험, 암 가족력, 체질량 지수, 알코올성 간질환, 당

뇨병, 고지혈증, 고혈압)을 통제했을 때, 폐경 전 여성은 폐경 후 여성보다 유방암 위험이 1.55-1.68배 높았고, 통계적으로 유의하였다.

넷째, 위암 발생 코호트의 알코올성 간질환 유무에서는 위암 발생 연관성을 나타내는 위험비와 95% 신뢰구간이 모델별에 따라 큰 차이가 존재하지 않았고, 다른 변수들(성별, 의료보험, 암 가족력, 흡연 상태, 당뇨병, 고지혈증, 고혈압)을 통제했을 때 알코올성 간질환이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 1.19-1.23배(모델 1, 2, 3, 4) 높았으며, 모두 통계적으로 유의한 결과를 보였다.

결론

시간 척도를 달리 선택함에 따라 위험비의 방향이 바뀌거나 통계적 유의성이 달라지는 차이가 존재한다는 것은 질병과 위험 요인의 연관성을 규명해내는 역학 코호트 연구에서 중요한 이슈가 될 수 있다.

왼쪽 절단된 코호트에서 연구 기간을 시간 척도로 사용하면 왼쪽 절단됨으로써 발생하는 편차 문제가 제기되므로 코호트 연구에서 콕스 비례위험모형을 사용하는 경우 도달 연령을 시간 척도로 하여 왼쪽 절단 분석하는 것을 제안하며, 적절한 시간 척도 선택을 위한 후속 연구가 계속 이루어져야 할 것이다.

핵심어: 왼쪽 절단, 시간 척도, 연구 기간, 도달 연령, 코호트, 연령 보정, 콕스 비례위험모형

Ⅰ. 서 론

1. 연구 배경 및 필요성

생존 분석은 관측 기간(Observation window)의 절단(Truncation)과 중도절단(Censoring) 관련 문제를 통계적으로 적절히 처리함으로써 생애 기간 동안 위치한 개인들을 효과적으로 분석 대상에 포함할 수 있는 장점이 있다(Van Hook et al., 2013). 역학 코호트 연구에서 생존 분석 시 흔히 사용하고 있는 모델은 콕스 비례위험모형으로(Cox D, 1972), 질병 발생과 잠재적인 위험요인 간의 관계를 평가하는데 많이 사용한다(Thiébaut AC et al., 2004). 임상 연구 및 종단 관찰 연구와 콕스 비례위험모형을 사용한 기존 논문들은 일반적으로 시간 척도(Time scale)를 연구 기간(Time-on study)으로 주로 사용해왔다(Korn E et al., 1997; Chalise P et al., 2009). 연구 기간(Time-on study)은 연구에 등록되어 사건(Event)이 발생할 때까지의 시간을 계산한 것을 말하며 <그림 1>로 설명할 수 있다. 이때 x축은 관측 기간(년)을, y축은 각각 5명의 연구 대상자를 나타낸다.

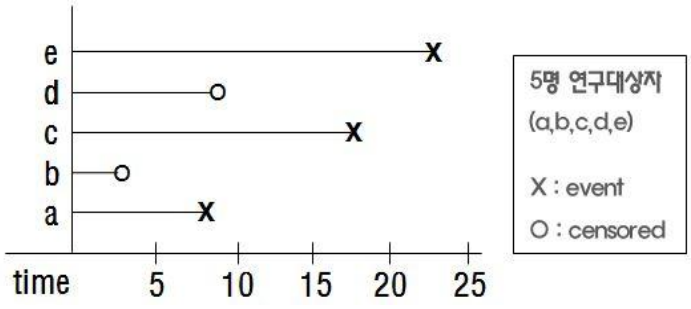


그림 1. Time-on study scale.

그러나, 시간 척도를 연구 기간으로 분석할 경우, 연구에 같은 시점에 등록된 대상자들이 동일한 관측 기간이 주어진다 하여 발생 위험이 모두 동일하다고 할 수 있는지에 대한 의문이 제기된다(Han, 2018). 또한, 나이는 보건학에서 만성질환 발생 위험에 상당한 영향을 주는 대표적인 혼란 변수(Confounding variable)로 연령 효과(Age effects)의 통제는 코호트 연구에서 중요한 문제이다. 이에 따라, 관측 기간을 연구 기간으로 사용한 경우, 연구에 등록된 시점의 연령을 보정(Age-adjustment)해줌으로써 분석하고 있다.

연구 기간을 시간 척도로 분석하면 때에 따라 편차(Bias)가 발생할 수도 있고, 발생하지 않을 수 있지만, 왼쪽 절단됨으로써 생기는 편차 문제는 해결하지 못할 수가 있다(Han, 2018). 즉, 코호트 연구에서 콕스 비례위험모형을 사용하는 경우, 단순히 연령만을 보정한다고 해서 발생하는 잠재적인 편차 문제를 해결하기 어렵다는 것이다. 이런 잠재적 편차를 해결하기 위한 방법의 하나로 도달 연령(Attained age)이라는 시간 척도를 적용해볼 수 있다<그림 2>. 도달 연령은 태어난 0세부터 사건이 발생할 때까지의 시간을 계산하는데, 사건이 발생하지 않았다면 마지막 관측 시점의 연령을 통상적으로 고려해준다. 그러나, 코호트 연구에서 대상자들은 각기 다른 시점과 연령에서 연구에 등록되기 때문에 <그림 3>과 같이 주로 왼쪽 절단(Left truncation)된 형태가 대부분 발생할 가능성이 있다(Lamarca R et al., 1998).

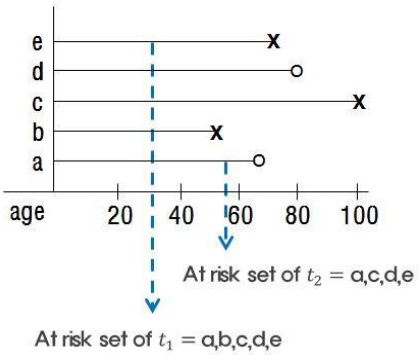


그림 2. Age scale.

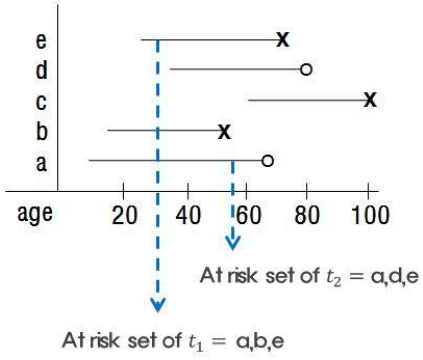


그림 3. Left truncated age scale.

<그림 2>와 <그림 3>에서는 <그림 1>과 달리, x축이 연령을 기준으로 적용된 것을 확인할 수 있다. 이때 사건이 발생할 수 있는 표본을 의미하는 위험집합(Risk set)을 시간 척도에 따라 비교해보면, 어느 한 시점을 t_i로 가정할 시 <그림 2>의 t_1시점에서는 a, b, c, d, e를 모두 포함하지만, Left truncated age scale인 <그림 3>에서는 아직 연구에 들어오지 못한 d, c를 제외하고 a, b, e만 포함된다. 마찬가지로 <그림 2>의 t_2시점에서는 사건이 발생한 b를 제외하고 a, c, d, e를 포함하지만, <그림 3>은 a, d, e만 포함된다. 즉, 시간 척도를 어떻게 선택하느냐에 따라 t_i시점 이후, 앞으로 사건이 발생할 가능성(Possibility)이 있는 사람들의 집합이 달라지고 있음을 확인할 수 있다.

절단(Truncation)은 일정한 관측 기간 안에 들어오지 못해 연구 대상에서 제외되는 것을 말한다. 이때, 왼쪽 절단은 왼쪽에서 발생하는 경우이다. 기존 선행연구들은 왼쪽 절단이 생존 추정치를 증가시키고, 공변량(Covariates)의 추정된 효과를 편향시킬 수 있는 가능성을 제기하고 있다(Fieberg J et al., 2009; Berg VD et al., 2011; Yang T and Aldrich HE, 2012).

그래서 Korn 등(1997)은 관찰 연구에서는 시간 척도를 도달 연령으로 사용하는 것이 더 적절하다고 주장하였다. 그러나, 다음 두 가지 경우일 때 연구 기간 분석과 도달 연령 분석은 유사한 결과를 보인다고 하였다. 첫째, 기저 위험 함수(Baseline hazard function)가 지수분포(Exponential distribution)를 따를 때, 즉 a는 연령, c와 Ψ가 상수라면 $\lambda_0(a) = cexp(\psi a)$를 의미한다. 둘째, 관심 있는 공변량 z와 코호트 진입 연령인 baseline age a_0가 독립적인 경우일 때 두 시간 척도 분석 결과의 편차는 줄어든다고 설명하였다.

하지만, Thiébaut 등(2004)은 Korn이 말한 두 번째 주장에 대해 상반된 결과를 보였다고 반박했다. 즉, 공변량과 코호트 진입 연령이 독립적일지라도 상당한 편차가 존재한다고 주장하였다. 반면, Pencina 등(2007)은 와이블 분포

(Weibull distribution)하에 코호트 진입 연령과 위험요인의 상관관계가 0(zero)일 때, 편차가 매우 가깝다는 것을 발견했다. 즉, 코호트 진입 연령과 위험요인이 독립이어도 연구 기간과 도달 연령 분석 간 유사한 결과를 보였다. Chalise 등(2009)은 Korn이 말한 두 가지 주장에 대해 모두 반박했다. 즉, 기저 위험함수(Baseline hazard function)가 지수분포(Exponential distribution)를 따를지라도, 코호트 진입 연령과 위험요인의 상관관계가 낮거나 또는 높더라도 두 시간 척도의 결과는 상당히 달랐다.

따라서 우리나라에서 콕스 비례위험모형에서 도달 연령을 시간 척도로 사용한 연구가 드물기 때문에 시간 척도를 달리 선택함에 발생하는 편차가 있을지 비교하는 것은 의미가 있고, 국내에서도 국민의 대표성을 지닌 국민건강보험공단의 코호트 자료를 이용해 왼쪽 절단된 데이터의 특성을 고려한 시간 척도를 적용해보는 시도가 필요하다고 생각된다.

2. 연구 목적

이 연구에서는 국민건강보험공단 검진코호트 DB(2002-2013년)를 이용하여 시간 척도(Time scale)를 연구 기간(Time-on study)과 도달 연령(Attained age)을 달리 선택함에 따라 발생하는 차이를 비교하고자 한다.

아래 두 가지 하위 연구주제에 대하여 구체적인 연구 목적은 다음과 같다.

가. 여성의 폐경 전·후 그룹에서 위험인자가 유방암 발생에 미치는 영향
나. 알코올성 간질환 유무에 따라 위암 발생에 미치는 영향

첫째, 연구 기간을 시간 척도로 분석한 콕스 비례위험모형과 도달 연령의 시간 척도에서 왼쪽 절단 분석한 콕스 비례위험모형 간 위험비와 베타 계수 값의 변화를 산출하고 비교한다.

둘째, 기저 위험 함수(Baseline hazard function)가 지수분포(Exponential distribution)를 따르는 여부와 코호트 진입 연령과 공변량 간 독립 여부를 파악하여, 각각 여부에 따라 차이가 달라지는 부분이 있는지 확인한다.

3. 선행 연구 고찰

가. 유방암 발생 현황 및 위험인자

국제암 보고서(World Cancer Report, 2014)에 따르면 암은 고소득 국가일수록 발생률이 높으며, 우리나라는 서유럽 및 북미와 함께 고소득 국가로 분류되어 암 발생률이 높은 국가에 속한다. 암 발생률은 1999년 이후에 2011년까지 연평균 약 3.8%씩 증가하다 2011년 이후 매년 3.0%씩 감소해왔으나, 유방암은 1999년 이후 발생률이 지속적인 증가 추세를 보인다. 2016년 기준 우리나라에서 유방암은 2005년 이후 지속적으로 1위를 유지했던 갑상선암을 제치고 여성 암 발생률 1위를 차지하며 유방암 연령표준화발생률(ASR, Age Standardized Rate)[1]은 10만 명당 약 62.5명이었다(한국중앙암등록본부, 2018).

유방암은 유방에 발생한 암세포로 이루어진 종괴를 의미하며, 일반적으로 유방의 유관과 유엽에서 발생하는 암을 일컫는다(한국유방암학회, 2018). 유방암도 다른 암과 마찬가지로 적절한 치료가 이뤄지지 않으면, 혈류 및 림프관을 따라 전신으로 전이되어 심각한 결과를 초래하기도 한다(한국유방암학회, 2018).

<그림 4>를 보면, 한국 여성 유방암 환자의 연령별 발생 빈도가 2000년부터 2016년까지 뒤집힌 V자 형태의 양상이 유지되고 있다(한국유방암학회, 2018). 2016년에 유방암이 가장 많이 발생한 연령군은 40대였으며, 이어 50대> 60대> 30대> 70대순의 발생 빈도를 보였다. 폐경 전·후를 기준으로 비교해보면, 2010년까지는 폐경 전 여성에서 유방암 비율이 폐경 후 여성보다 높았으나, 2011년부터 폐경 전 여성의 유방암 환자 비율은 점차 줄어들고, 폐경

1) 한국중앙암등록본부의 국가암등록사업 연례 보고서(2018)에 따르면, 연령표준화발생률(ASR)은 각 연령군에 해당하는 표준인구의 비율을 가중치로 주어 산출한 가중 평균 발생률을 의미하며, 해당 자료의 표준인구는 우리나라 2000년 주민등록연앙인구를 사용함. 식으로 표현하면, 다음과 같음. 연령표준화발생률=Σ(연령군별발생률×표준인구의 연령별 인구)/표준인구

후 여성의 비율은 증가하는 양상을 보였다(한국유방암학회, 2018). 유방암 환자의 중간 나이는 2000년에는 46세였으나 2011년 이후에는 50세 이상, 2016년에는 51세로 중간 나이가 점차 높아지고 있다(Kang et al., 2018) <그림 5>.

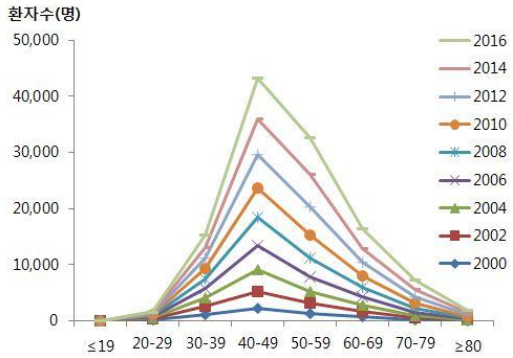


그림4. 연도별 연령별 유방암 환자 수(여자).
(보건복지부, 「암등록통계」)

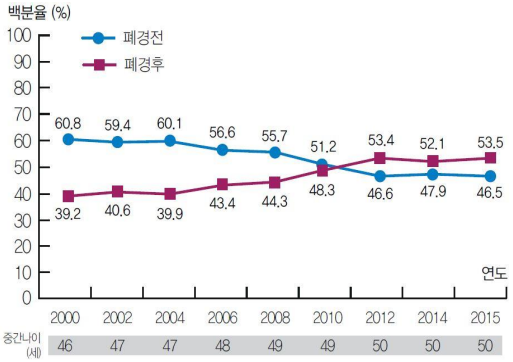


그림5. 폐경 전후에 따른 유방암 발생 빈도.
(한국유방암학회 2018)

유방암은 내인성 및 외인성 호르몬 등 성호르몬 요인, 비만, 신체활동 등 생활습관 요인과 알코올, 흡연, 방사능 노출 등 환경 요인, 유전적 요인 등 매우 다양한 요인들이 복잡하게 얽혀 있는 질병으로 알려져 있다(박수경, 2019). 현재까지 유방암 발생 원인이 뚜렷히 규명된 상태는 아니나, Yoo 등(2002)과 Park 등(2016) 한국 여성의 유방암을 대상으로 한 연구를 통해 입증된 위험요인들이 존재한다. 위험도를 증가시키는 요인에는 성호르몬 요인과 관련된 이른 초경, 늦은 폐경, 폐경 후 여성에서의 비만, 유방암 가족력 등이 있다. 특히 늦은 연령의 첫 만삭 임신에서 일반 인구 기여 위험도가 8.0% 정도였다. 위험도를 감소시키는 요인으로는 모유 수유, 운동, 야채와 과일 섭취 등이 알려져 있다.

그러나, 과거 유방암의 여러 위험요인들은 한국인에서의 위험요인과 유방암 연관성에 대한 결과가 서양 여성에서의 결과와 서로 다르게 관측되는 일관적이지 않은 결과를 보였다. 그 이유로는 기존 선행연구들의 한국인 대상 유방암

역학연구 수가 상대적으로 적었고 대부분이 소규모 연구로 한 병원 기반 환자 대조군 연구였기 때문인 것으로 파악된다(박수경, 2019).

나. 관련 선행연구 고찰

폐경 전·후에 따라 각 위험요인과 유방암 간의 관련성을 보려 하는 국내 선행연구들은 진단 시 연령을 통제 및 층화하거나 또는 초경 및 폐경의 시작 연령을 층화 분석한 연구들이 많았다. 도달 연령을 적용한 연구는 찾기 어려웠으며, 왼쪽 절단(Left truncation)에 대한 가능성을 고려하지 않은 연구들이 대부분이었다(Park et al., 2014; Park et al., 2016; Jeong et al., 2017).

오현경(2010)은 환자-대조군 연구에서 연령 및 폐경에 따른 효과 조정 현상을 분석하기 위해 폐경 유무로 나누어 층화 분석을 하였고, 진단 시 연령을 교란 변수로 두어 분석하였다. 최근 2년간 체중 감소는 폐경 전 여성에서 유방암의 사망 위험도를 증가시킬 수 있는 결과가 나왔지만 출산력, 모유 수유 기간, 초경 나이, 가족력은 폐경 전과 폐경 후 여성 모두에서 유방암의 사망 위험도와 유의한 관련성을 보이지 않았다. 이때 폐경의 기준은 1998년부터 2004년까지 삼성서울병원에서 모집된 환자들로, 1년 이상 월경이 나오지 않을 경우 폐경으로 간주하여 폐경 전·후로 구분하였다.

임선미 등(2011)은 국민건강보험공단의 1990-1992년 공무원 및 사립학교 교직원을 대상으로 구축한 건강검진 코호트를 가지고, 폐경 시작 평균연령을 기준으로 저연령군과 고연령군으로 나누어 분석하였다. 연구 시작 시점에서 체질량 지수의 평균과 유방암 발생의 관련성은 고연령군에서만 경향성과 위험 증가가 관찰되었고, 저연령군에서는 발생 위험이 낮아지기는 했으나 통계적으로 유의하지는 않은 결과를 보였다. 이때 국민건강보험공단의 건강검진 설문지에는 폐경 여부 조사항목이 부재하였으므로 여성의 건강통계 결과에서 보고한 우리나라 여성의 평균 폐경 시작 연령 47.91세를 적용하여 분석하였다(Ministry of

Health and Welfare, 2004),

Lee 등(2018)은 우리나라 국민건강보험공단 자료를 이용한 코호트 연구로 연령을 포함한 독립변수들을 통제한 콕스 비례위험모형으로 분석하였다. 이때 폐경 후 여성에서는 체질량 지수를 통제하거나, 하지 않는 모델에서 허리둘레는 모두 유방암 위험과 밀접한 관련이 있었지만, 폐경 전 여성에서는 체질량 지수를 통제한 모델에서만 통계적으로 유의했다(Lee et al., 2018). 이렇듯 종단관찰연구와 콕스 비례위험모형을 사용한 기존 국내 논문들은 일반적으로 시간 척도를 연구 기간으로 하되, 연령을 보정한 연구들이 대부분이었다. 이와 달리, 국외 연구사례 중 Cheung 등(2003)은 유방암 진단받은 환자에서 생존을 다른 시간 척도를 적용함으로써 비교하였는데, 콕스 비례위험모형에서 도달 연령을 시간 척도로 했을 때는 유방암을 진단받은 젊은 여성에서 높은 사망률과 관련이 있었지만, 연구 기간을 시간 척도로 하면 비교위험도와 신뢰구간의 방향이 정반대로 바뀌는 결과를 보였다.

다. 위암의 발생 현황 및 위험인자

2018년 기준 전 세계적으로 위암은 다섯 번째로 흔히 발생하는 암으로, 전체 암 사망의 8.2%를 차지하여 세 번째로 흔한 사망원인이다(Bray et al., 2018). 특히 우리나라는 급격한 고령화, 빠른 경제성장, 생활습관의 서구화, 의학기술의 발전 등 사회경제적, 문화적, 보건학적으로 많은 변화가 있었기에 위암 발생률, 사망률, 생존율에 영향을 주었고 관련 위험요인을 확인할 기회가 되기도 했다(고광필, 2019).

위암은 최근 들어 전 세계적으로 감소하는 추세이기는 하나(Torre et al., 2016), 2018년 중앙암등록본부 자료에 의하면 2016년 기준 갑상선암을 제외한 암 중 우리나라 암 발생률 1위를 차지하는 암이다. 2016년 위암 발생자 수는 30,504명으로 연령표준화 위암 발생률은 인구 10만 명당 35.4명이었다(한국중앙암등록본부, 2018).

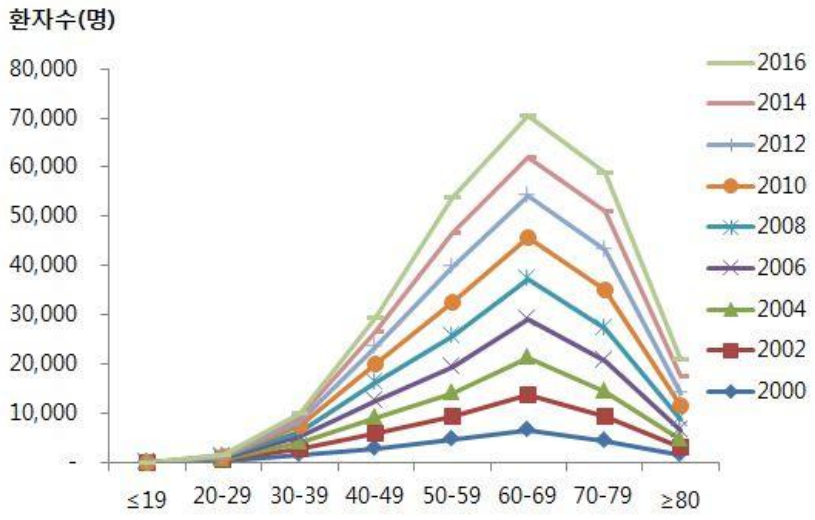


그림 6. 연도별 연령별 위암 환자 수.
(보건복지부, 「암등록통계」)

우리나라 위암 환자의 연령별 발생 빈도는 2000년부터 2016년까지 연령이 증가함에 따라 발생률이 증가하고 있는 양상을 보이다가 고령인 70대 구간에서부터 오히려 감소하고 있다. 2016년에는 60대가 위암이 가장 많이 발생한 연령군이었으며, 60대> 70대> 50대> 80대 이상> 40대 순의 발생 빈도를 보였다<그림 6>. 위암은 지역, 인종, 성별, 사회경제적 수준 등에 따라 현저한 발생과 사망의 차이를 보이는 암으로 약 70% 정도가 개발도상국에서 일어나며, 약 절반이 동아시아에서 발생한다(송민교, 이휘원, 강대희, 2015).

위암의 위험요인에는 유전적 소인뿐만 아니라, 헬리코박터파일로리(Helicobacter pylori) 감염과 식습관 등을 포함한 환경적인 요인, 흡연 및 음주 등의 행동적 요인, 소금에 절인 음식 섭취 등 식이 요인이 매우 크게 작용하고 있는 것으로 알려져 있다(Correa, 1999; Tsugane and Sasazuki, 2007; Guggenheim and Shah, 2013). 또한, 위암은 가족 중에 위암이 있을 경우에 위험도가 약 2-3배(최대 9.9배) 많이 발생하는 결과가 있었으며(Zanghieri et al., 1990; Yaghoobi et al., 2010), 국내에서는 위암 직계가족에서 헬리코박터파일로리에 감염된 경우 위암 발생률이 5.3배 증가하는 것으로 보고했다(Shin et al., 2010).

흡연은 국제암연구소(IARC, International Agency for Research on Cancer)에서 발암 1군으로 분류한 위암 발생에 근거가 있는 환경인자로(송민교, 이휘원, 강대희, 2015). 역학연구에서는 흡연과 위암 발생 간의 관련성이 있는 일관된 연구 결과를 보인다(고광필, 2019). 한 메타분석 연구에서는 흡연자의 경우 비흡연자에 비해 1.28배(95% 신뢰구간 : 1.17-1.41) 위험이 증가하였으며 분문부 위암과 비분문부 위암에서 모두 위험이 증가하였다(Ferro et al., 2018). 일본의 Ladeiras-Lopes 등(2008) 연구에서는 흡연 시작 연령이 빠를수록, 하루 흡연량이 많을수록 그 위험도가 더 높아지는 양-반응 효과가 있는 결과를 보이기도 했다. 이밖에 제2형 당뇨병도 암의 위험요인으로, 공복 혈당 수치가 증가함에 따라 암 사망률이 증가하였으며, 특히 남성의 경우 위암 등 암 발생

위험을 증가시켰다(Jee et al., 2005; Vigneri et al, 2009; 정한영, 이숙향, 2019). 음주의 경우 국제암연구소와 세계암연구재단(WCRF, World Cancer Research Fund)의 보고서에서 하루 3잔 이상의 음주는 위암 발병에 강한 근거가 있는 위험요인으로 규정하였다(World Cancer Research Fund, 2018).

라. 관련 선행연구 고찰

장기간의 과다한 음주는 알코올성 지방간, 알코올성 간염, 간경변증 등 알코올성 간질환을 초래하기도 하지만, 국내에서는 알코올성 간질환과 위암 발생의 관계에 대한 대규모 연구를 찾아보기 어려웠다.

대신 많은 위암의 위험요인 중 음주와의 연관성에 대한 연구는 많이 진행되어 왔고, 유방암과 비슷하게 위암에서도 종단 관찰연구와 콕스 비례위험모형을 사용한 기존 선행연구들은 연구 기간을 시간 척도를 적용하면서, 연령을 보정한 논문들이 주를 이루었다(Sung et al., 2007; Yi et al., 2010; Jung et al., 2012; Tramacere et al., 2012; Choi et al., 2017; Minami et al., 2018). 이때 알코올 섭취량과 위암과의 연관성은 전 세계적으로 연구마다 그 결과가 다르게 보고되기도 했다. 알코올 섭취량과 위암과의 연관성이 유의한 결과가 있는 반면(Sung et al., 2007; Jung et al., 2012; Choi et al., 2017), 유의하지 않는 결과를 보이거나(Minami et al., 2018), 과도 음주자의 경우 비음주자에 비해 위암 발생 위험이 유의하게 증가하거나(Tramacere et al., 2012), 성별에 따라서 다른 결과를 보이기도 했다(Yi et al., 2010).

Ⅱ. 이론적 배경

1. 왼쪽 절단(Left truncation)

먼저, 관찰 연구에서 흔히 발생할 수 있는 자료의 형태인 절단(Truncation)은 일정한 관측 기간(Observation window) 안에 들어오지 못해 연구 대상에서 제외되는 것을 말한다. 즉, 절단 자료(Truncation data)란 분석할 자료가 전체 집단의 어떤 조건을 만족하는 부분 집단만으로 구성된 자료를 의미한다. 또한, 주어진 조건보다 큰 생존 시간만이 분석 대상이 되는 경우도 해당한다. 예를 들어, 60세 이상만을 연구 대상자로 모집한다면, 60세 이전의 사람들은 코호트에 들어올 수 없게 될 때를 말한다.

이때, 절단은 왼쪽 절단(Left truncation)과 오른쪽 절단(Right truncation) 두 가지의 경우가 있다. <그림 7>에서 연구 시작 시점을 (a), 끝나는 시점을 (b), 총 관측 기간을 (c)로 본다면, 왼쪽 절단(Left truncation)은 연구 시작 전에 사망하였거나, 사건이 발생하여 연구에 포함되지 않은 B를 의미한다. 어느 관측 기간을 (Y_L, Y_R)로, 사건 발생을 X 라고 가정해보자. 왼쪽 절단 자료가 존재한다면, 사건이 관측 기간보다 전에 발생하여 즉, $Y_L < X$ 되어, 그 자료는 관측되지 않는다(Klein, 2003). 개체가 사망하거나 절단(Censored)될 때까지 관측하기 때문에 왼쪽 절단을 *Delayed entry time* 또는 *Delayed entry* 라고 부르기도 한다(Klein, 2003).

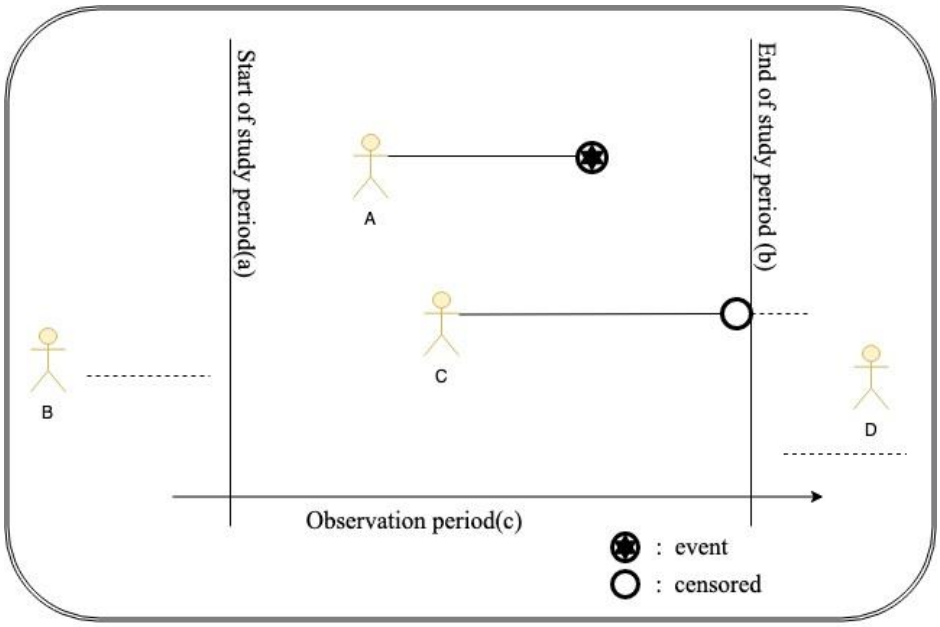


그림 7. 왼쪽 절단(Left truncation)과 오른쪽 절단(Right truncation).

<그림 7>에서 D는 오른쪽 절단(Right truncation)의 경우에 해당한다. 오른쪽 절단은 단면 조사 연구의 샘플링(Sampling), 후향성 연구의 샘플링(Sampling)과 같이 특정한 표본 추출 방식에서 자연스럽게 일어나며 특히, 질병의 잠복기를 다루는 임상 연구에서 때때로 발생한다(Wang, 1999; Bilker and Wang, 1996; Chi, Tsai and Chiang, 2007). 에이즈(AIDS, 후천성 면역결핍증후군)를 예로 들어보면, 에이즈 잠복기는 HIV 바이러스 감염에서 에이즈 진단까지의 시간으로 정의된다(Jiang, 2011). 이때 오염된 혈액이 수혈을 통해 전염되면, HIV 바이러스의 근원이 될 수 있다. Y_L를 에이즈 걸린 피를 수혈 받은 시점, Y_R을 관측이 끝나는 시점, L을 HIV 바이러스의 잠복기라 가정해 보자. 관측이 끝나기 전에 에이즈를 진단받은 환자들은 표본에 포함되며 즉, Y_L+L≤ Y_R 이거나, L≤ Y_R− Y_L이 된다(Jiang, 2011). 그러나, 에이즈가 발견되

기 전까지인 잠복기 단계에서 HIV에 감염됐는지, 안됐는지 알 수 없으면 표본에 포함되지 않게 되기 때문에 오른쪽 절단되는 것이다. 즉 관측이 종료된 후에 에이즈에 걸린 환자들은 모집단에는 속하지만, 관측 시점까지 에이즈에 걸리지 않은 사람들로 간주하므로 표본에 속하지 않게 되는 것이다(Chi, Tsai and Chiang, 2007).

그밖에 A는 연구 대상자로 연구에 포함되어 관측 기간 안에 사건이 발생하는 경우를 의미하고, C는 사망 또는 관측 기간 안에 사건이 발생하지 않은 등의 이유로 절단(Censored)된 경우이다.

2. 콕스 비례위험모형(Cox's Proportional Hazard Model)

콕스 비례위험모형은 Cox(1972)가 제안한 모형으로 생존 함수가 지수함수(Exponential function)를 따른다는 것과 두 군의 위험 함수의 비가 연구 기간 동안 일정하게 유지된다는 두 가지 중요한 가정에서 출발한다(In JY and Lee DK, 2018). 생존 시간 T의 분포에 대한 가정을 필요로 하지 않고, T와 공변량의 관계를 다음과 같은 위험 함수(Hazard function)로 정의한다.

$$h(t|X) = \lim_{\triangle t \to 0} \frac{P(t \le T \le t + \triangle t | T \ge t)}{\triangle t} = h_0(t)\exp(\beta X)$$

이때, $h_0(t)$는 특정 t 시점에서 모든 공변량 값들이 0일 때의 순간 위험률로 분포 및 형태에 대한 가정이 주어져 있지 않은 기저 위험 함수(Baseline hazard function)를 말한다.

누적 위험 함수(Cumulative hazard function)를 $H(t) = \int_0^t h(u)du$ 라고 하면 (Breslow, 1974), 생존 함수 $Q(t)$는 다음과 같이 표현할 수 있다.

$$Q(t) = \exp\left[-\int_0^t h(u)du\right] = \exp[-H(t)]$$

콕스 비례위험모형은 i번째 사람의 위험률과 j번째 사람의 위험률의 비가 시간에 대한 일정함을 가정한다. i번째 사람과 j번째 사람의 위험 함수의 비는 다음과 같이 표현할 수 있다.

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}{h_0(t)\exp(\beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_k x_{jk})}$$
$$= \exp\{\beta_1(x_{i1} - x_{j1}) + \beta_2(x_{i2} - x_{j2}) + \cdots + \beta_k(x_{ik} - x_{jk})\}$$

즉, 기저 위험 함수는 상쇄되어 없어지므로, 위험비는 하나의 상수(위험비=λ)가 되며, 이를 비례위험도(Hazard Ratio)라고 한다.

3. 생존 분석의 시간 척도 선택

이 연구에서 적용해보고자 하는 모형은 총 4개이다.

모델 1은 연령을 보정하지 않고, 연구 기간을 시간 척도로 한 모형이다. 이 때 T는 연구 기간을 의미한다.

$$h_t(t|a_0,x) = h_{0T}(t)e^{\gamma x} \qquad (1)$$

모델 2는 연령을 보정하고, 시간 척도를 연구 기간으로 사용한 함수를 나타내었으며 이는 역학 연구에서 일반적으로 가장 많이 사용한다. T는 연구 기간을, a_0은 각 대상자의 연구 시작 시점의 연령을 의미한다.

$$h_t(t|a_0,x) = h_{0T}(t)e^{\xi a_0 + \gamma x} \qquad (2)$$

모델 3은 시간 척도를 도달 연령으로 사용하고, 출생연도를 구간으로 층화(Stratified)한 모형이다. 각 $B_1, B_2, \ldots, B_i$로 i을 출생연도의 구간 값으로 넣은 것을 의미한다. 이는 Korn 등(1997)이 가장 추천한 모형으로 연령효과(Age effects), 코호트효과(Cohort effects)뿐만 아니라, 기간 효과(Period effects)까지 보정할 수 있는 모델이다. 이 연구에서는 5년 단위로 구간을 주어 분석하였다.

$$h_A(a|b_0 \in B_i,x) = h_{oiA}(a)e^{\beta x} \qquad (3)$$

모델 4는 시간 척도를 도달 연령으로 하고, 코호트가 등록되는 시점을 왼쪽 절단시킨 것을 의미하며, 이를 수식으로 표현하면 (4)와 같다. T는 사건이 발생할 때의 연령이고, 사건이 발생하지 않은 경우는 마지막 관측 시점의 연령을 의미한다. 이때 a는 도달 연령이고, a_0은 각 대상자의 코호트에 들어온 시점의 연령으로 왼쪽 절단된 시점을 의미한다(Commenges D, Letenneur L and Joly PJAjoe, 1997; Han, 2018).

$$h_A(a|a_0, x) = h_0(a)e^{LT_x}, \quad \text{a} > a_0 \qquad (4)$$

Ⅲ. 폐경 전·후 여성에서 위험인자가 유방암 발생에 미치는 영향

1. 연구 방법

가. 연구 자료

국민건강보험공단(이하 건강보험공단)의 건강검진코호트 DB(이하 검진코호트 DB)를 이용하여 폐경 전·후에 따른 위험인자가 유방암 발생에 미치는 위험을 파악하고 시간 척도를 연구 기간 대비 도달 연령에 따른 차이를 비교하고자 한다. 검진코호트 DB의 2002년부터 2013년까지 12년 동안 건강보험공단에 청구된 데이터를 활용했다.

검진코호트는 건강검진 수검자의 의료이용 및 건강결과 분석을 위해 구축된 연구용 DB로, 우리나라 40세 이상 국민의 대표성을 지닌 코호트 자료이다. 검진코호트의 모집단은 2002-2003년 일반 건강검진 수검자 중 2002년 12월 말 기준 중복 수검자를 제외한 40세부터 79세의 건강보험 자격 유지자 515만 명이다. 연구에 사용된 표본대상자는 514,866명으로 모집단의 10% 단순무작위추출로 선정되었다. 검진코호트 구축 DB는 자격 DB, 진료 DB, 건강검진 DB, 요양기관 DB로 구성되어 있다. 구축 내용으로는 사회·경제적 자격 변수(장애 및 사망 포함), 병·의원 이용내역 및 건강검진 정보, 요양기관 현황 등을 포함하고 있으며, 각 DB에 대한 구성은 표 1과 같다. 이 연구에서는 자격 DB와 진료 DB에서 명세서(20t)와 상병내역(40t), 그리고 건강검진 DB를 이용하였다.

표 1. 검진코호트 DB구성[2)]

구분	구성내용	
자격 DB	인구사회학적 정보(성, 연령, 거주지역), 사망관련 정보(사망일자, 사망원인), 건강보장 유형(건강보험 가입자 구분/의료급여), 사회경제적 수준 관련 자료	
진료 DB	대상자가 요양기관에 방문하여 진료 받은 내역을 요양기관이 청구하여 심사가 결정된 요양급여 내역 자료 ※ 의과_보건기관(T1), 치과_한방(T2), 약국(T3) 자료에 대한 명세서(20t), 진료내역(30t), 상병내역(40t), 처방전교부상세내역(60t)의 세부 DB로 구성	
	명세서(20t)	요양기관(병원, 의원, 약국 등)에서 진료, 조제 등의 의료서비스를 제공하고 청구방법에 따라 작성한 청구단위 명세서 공통내역 자료 ※ 주상병과 부상병 1순위만 포함
	진료내역(30t)	명세서에 따른 진료행위, 의약품, 치료재료 등의 행위별 상세 진료내역 및 원내처방내역 자료
	상병내역(40t)	주상병과 부상병 1순위뿐만 아니라, 주상병, 부상병 및 추가 부상병이 모두 포함된 자료
	처방전교부 상세내역(60t)	처방전교부건별 원외처방내역 관련 상세 자료
건강검진 DB	건강검진 주요 결과 및 문진에 의한 생활습관 및 행태관련 자료	
요양기관 DB	요양기관 종류, 지역(시도)별 현황, CT 및 MRI, PET 등 장비, 시설, 인력관련 자료	

2) 국민건강보험공단, 건강검진코호트DB 사용자 매뉴얼. 2016

표 2의 상병 관련 변수는 모두 한국표준질병·사인분류(KCD)를 참조하여 코딩되었다. 사망원인 정보는 통계청 자료와 연계된 항목으로, 사망원인 1은 일반적인 사망(A00-R99)만 부여하며, 사망원인 1(DTH_CODE1)에 기재된 사망원인이 S00-T98인 경우에는 상세원인으로 사망원인 2에 중분류 코드(V01-Y98)를 부여한다.

질병의 진단명은 진료 DB의 명세서 일반내역(20t)과 수검자 상병내역(40t)에서 확인할 수 있다. 명세서 일반내역(20t)에는 주상병과 부상병 1순위만을 포함하고 있으나 수검자 상병내역(40t)에서는 주상병과 부상병을 포함한 청구명세서의 모든 상병이 포함되어 있다.

표 2. 상병 관련 변수 설명

구분	검진 DB		포함범위	해당 변수명 (마스터코드)	설명
사망원인	자격 DB		사망원인	사망원인 1 (DTH_CODE1)	KCD코드에 따른 사망원인 코드
				사망원인 2 (DTH_CODE1)	사망원인이 S00-T98인 경우 상세 원인 기재(V01-Y98)
진단명	진료 DB	명세서 (20t)	주상병, 부상병 1순위	주상병 (MAIN_SICK)	진료기간 중 치료나 검사 등에 대한 환자의 요구가 가장 컸던 상병
				부상병 (SUB_SICK)	진료기간 중 주상병과 함께 있었거나 발생된 상병으로 환자 진료에 영향을 주었던 상병
		상병내역 (40t)	모든 상병	상병기호 (SICK_SYM)	주상병과 부상병을 포함한 청구명세서의 모든 상병

나. 연구 대상

연구 대상은 2002년 검진코호트 자격 유지자 기준 514,866명을 대상으로 추출을 시작하였다. 남자를 제외하고, 여자에서 1차 일반 건강검진을 수검한 235,741명을 연구 대상자로 선정하였다.

일반건강검진의 특성상, 2년에 1회 주기로 수검을 받기 때문에 2002년에 건강검진을 받지 않았던 대상자는 2003년에 1차로 받은 일반 건강검진 DB를 활용했다. 반면, 매년 수검 받는 비사무직 직장가입자처럼, 2002년과 2003년 모두 수검을 받았던 대상자는 2002년에 받은 검진 결과를 기준으로 활용하였다.

연구 기간 동안 처음 진단된 환자들만 대상으로 하기 위해 2년의 워시아웃(Wash-out) 기간을 두어 2002-2003년에 사망한 대상자(323명)를 제외하였다. 또한, 2004년 이전 2년 동안 유방암으로 청구된 이력이 있는 대상자(1,047명)는 분석에서 제외하였다. 그리고 일반 건강검진 1차 문진 항목에서 과거력 항목에 암(1,332명)을 체크한 사람도 순차적으로 제외하였다. 마지막으로 건강검진 1차 문진 항목 중 체질량 지수와 암 가족력 결측치 인원(24,728명)을 제외하고 208,257명을 최종 연구 대상자로 선정하였다. <그림 8>

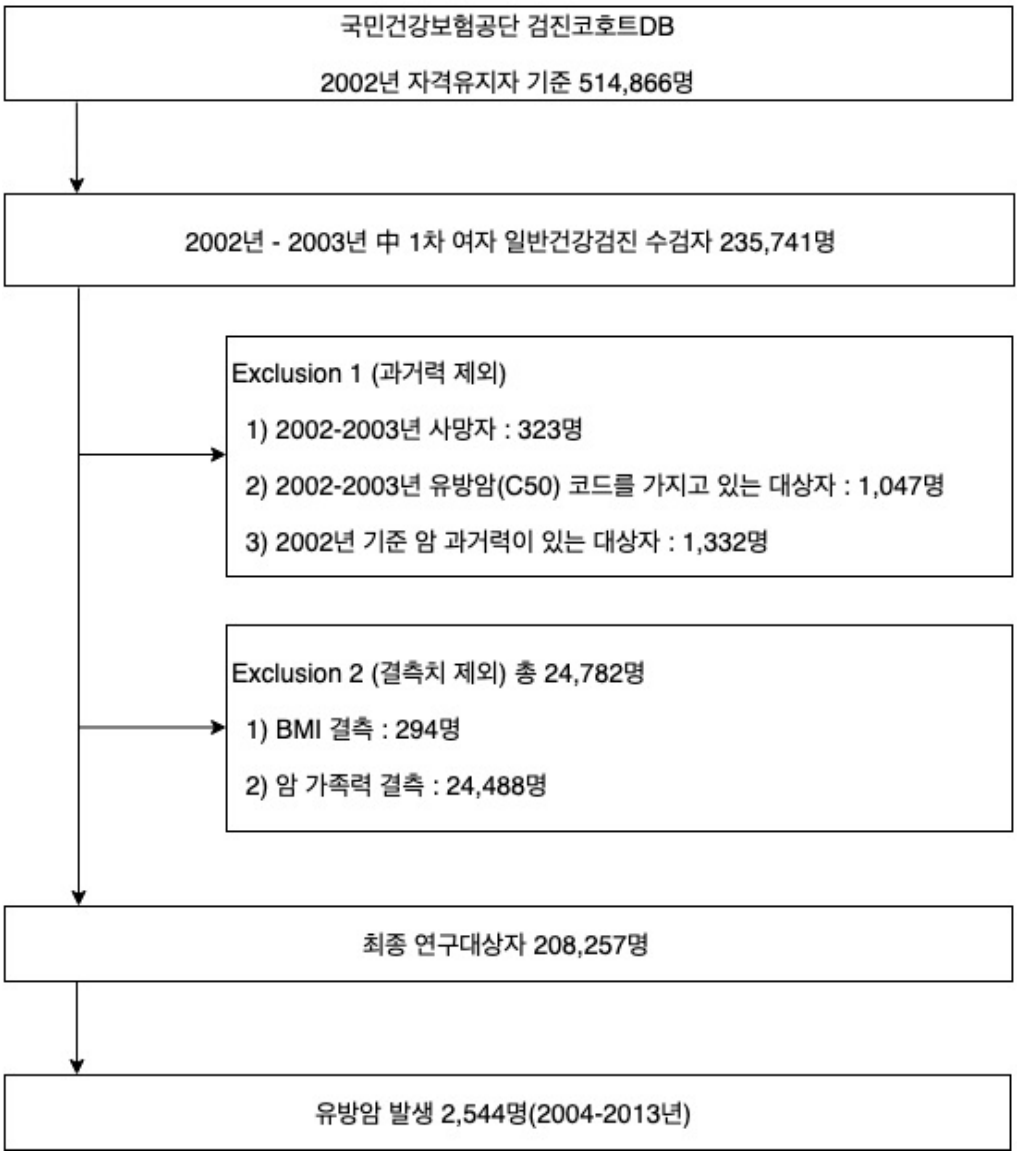


그림 8. 유방암 발생 코호트의 연구 대상자 선정.

다. 변수 정의

1) 유방암(Brest Cancer)

연구 대상자의 유방암 진단은 건강보험공단 검진코호트 진료 DB의 명세서(20t)를 확인하였으며, 상병코드는 한국표준질병 사인분류(KCD) 중 C50의 코드로 2004년부터 2013년 이내에 주상병 또는 부상병 1순위로 진단받은 사람으로 사용하였다.

2) 폐경(Menopause)

건강보험공단의 건강검진 설문지에는 폐경 여부를 확인할 수 있는 조사 문항이 없으므로 명확한 폐경 여부를 알 수 없다. 대안으로 한국 여성의 건강통계 결과에서 보고한 우리나라 여성의 폐경 시작 연령 평균 47.91세를 적용하여(Ministry of Health and Welfare, 2004), 폐경 전(47세 미만 군, Pre-menopause)과 폐경 후(47세 이상 군, Post-menopause)로 구분하였다(임선미 등, 2011).

3) 연령(Age)

검진코호트 자격 DB의 2002년 기준인 연구가 시작되는 시점에서 연구 대상자의 연령을 적용하였다. 연구 기간과 도달 연령 분석을 비교해보기 위해 연령 구간을 따로 나누지 않고 코호트에 구축된 40세에서 79세까지의 연령을 그대로 사용하였다.

4) 의료보험 종류(Medical Insurance Type)

자격 DB에서 지역가입자 및 직장가입자인 일반 건강보험가입자(NHI)와 의료급여수급자(medical aid) 두 군으로 나누어 구성하였다.

5) 암 가족력(Cancer Famliy History)

일반 건강검진 1차 문진에 있는 암 가족력 변수를 활용하였으며, 가족력이 없는 군과 있는 군으로 나누어 구성하였다. 단, 2002년에 받은 건강검진을 기준으로 적용하되, 2002년에 검진을 받지 않은 대상자는 2003년에 1차로 받은 일반 건강검진 DB를 활용하고, 2002년과 2003년 모두 수검을 받았던 대상자는 2002년에 받은 건강검진 DB를 적용하였다.

6) 체질량 지수(Body Mass Index, BMI)

일반 건강검진 1차 문진에 있는 체질량 지수(BMI)를 활용했으며, 세계보건기구가 발표한 아시아 기준 비만(WHO expert consultation, 2004)에 따라 underweight(<18.5 kg/㎡), normal(18.5-22.9 kg/㎡), over-weight(23.0-24.9 kg/㎡), obese class Ⅰ(25.0-29.9 kg/㎡), obese class Ⅱ(≥30.0 kg/㎡)인 5가지 그룹으로 나누어 구성하였다. 단, 2002년에 받은 건강검진을 기준으로 적용하되, 2002년에 검진을 받지 않은 대상자는 2003년에 1차로 받은 일반 건강검진 DB를 활용하고, 2002년과 2003년 모두 수검을 받았던 대상자는 2002년에 받은 건강검진 DB를 적용하였다.

7) 알코올성 간질환(Alcoholic Liver Disease)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 K70의 코드로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 알코올성 간질환으로 정의하였다.

8) 당뇨병(Diabetes Mellitus, DM)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 제1형 당뇨를 제외한 E11, E12, E13, E14의 코드로

청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 당뇨병으로 정의하였다.

9) 고지혈증(Hyperlipemia)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 지질단백질대사장애 및 기타 지질증(KCD-10, E78)으로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 고지혈증으로 정의하였다.

10) 고혈압(Hypertension)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 본태성(일차성) 고혈압(KCD-10, I10), 고혈압성 심장병(KCD-10, I11), 고혈압성 신장질환(KCD-10, I12), 고혈압성 심장 및 신장질환(KCD-10, I13), 이차성 고혈압(KCD-10, I15)으로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 고혈압으로 정의하였다.

표 3. 유방암 발생 코호트 변수 정의

변수		정의
독립 변수	폐경	2002년 연령 기준 폐경 전(47세 미만군, Pre-menopause), 폐경 후(47세 이상군, Post-menopause)로 구분
혼란 변수	연령	2002년 기준 검진코호트 등록(Enroll)된 때 연령
	의료보험	건강보험가입자(NHI)와 의료급여수급자(medical aid)
	암 가족력	일반 건강검진 1차 결과 암 가족력 없음, 있음
	체질량 지수	일반 건강검진 1차 결과 18.5kg/㎡ 미만, 18.5-22.9kg/㎡, 23.0-24.9kg/㎡, 25.0-29.9kg/㎡, 30.0kg/㎡ 이상으로 구분
	알코올성 간질환	2002-2003년 사이 주상병 또는 부상병(K70)으로 청구된 자
	당뇨병	2002-2003년 사이 주상병 또는 부상병(E11, E12, E13, E14)으로 청구된 자
	고지혈증	2002-2003년 사이 주상병 또는 부상병(E78)으로 청구된 자
	고혈압	2002-2003년 사이 주상병 또는 부상병(I10-I13, I15)으로 청구된 자
종속 변수	유방암	연구 기간(2004-2013년)사이에 주상병 또는 부상병 1순위(C50)로 최초 진단받은 환자

라. Time-to-event 정의

건강보험공단 청구내역 대상자 중 연구 기간(2004-2013년) 사이에 주상병 또는 부상병으로 유방암(C50)을 진료받은 청구내역이 있는 경우를 유방암 발생자로 하였다. 또한, 유방암 청구내역은 없지만, 사망원인이 유방암인 경우도 암 발생자로 가정하였다.

Time-to-event는 검진코호트 DB의 시작일인 2002년 1월 1일을 기준으로 유방암으로 청구된 첫 번째 요양 개시일까지 시간으로 정의하였다. 유방암 청구내역은 없지만, 사망원인이 유방암인 대상자는 사망일을 Time-to-event로 정의했다. 2013년까지 유방암으로 청구된 건이 없는 대상자는 사망 여부에 따라 아래 표 4와 같이 중도절단(Censored) 하였다.

연구 기간 대비 도달 연령을 비교 분석하기 위해 다음과 같이 도달 연령을 정의하였다. 유방암 발생자는 발생한 때의 연령을, 유방암이 발생하지 않은 생존자는 연구 기간이 종료되는 2013년 시 연령을 고려해주었다. 또한, 유방암 청구내역은 없으나 사망원인이 유방암이었던 대상자와 유방암이 발생하지 않은 사망자는 사망 시 연령으로 정의하였다.

표 4. 유방암 발생 코호트 Time-to-event 정의

사건발생	정의	사건이 발생할 때까지의 시간	도달 연령
유방암 발생	유방암 발생자	유방암으로 청구된 첫 번째 요양 개시일	유방암이 발생한 연령
	유방암 청구내역은 없지만, 사망원인이 유방암인 대상자	사망일	사망 시 연령
유방암 발생하지 않음	사망자	사망일	사망 시 연령
	사망하지 않은 자(생존자)	연구 기간이 종료되는 일자 (2013년 12월 31일)	연구 기간이 종료된 2013년 시 연령

마. 분석 방법

관측 기간에 여성의 폐경 전·후에 따라 위험 요인이 유방암 발생에 영향을 미치는지 확인하고자, 유방암 발생에 영향을 미칠 수 있는 가능성을 차단하기 위해 2년의 워시아웃(Wash-out) 기간을 두어 2004년 이전 2년 동안 유방암 및 암 과거력을 가지고 있는 대상자는 연구에서 제외하였다.

이 연구에서는 일반적으로 생존 분석 시 흔히 사용하는 연구 기간 시간 척도를 적용한 콕스 비례위험모형과 Korn 등(1997)이 제안한 도달 연령 시간 척도를 적용한 콕스 비례위험모형을 비교함으로써 발생하는 차이를 확인하였다.

세부적인 분석은 다음과 같다. 첫째, 연구 대상자의 일반적 특성을 파악하기 위해 폐경 전·후에 따라 카이스퀘어 검정(Chi-square test)을 이용하여 빈도와 백분율을 구하고, 필요에 따라 경향성 검정을 하였다.

둘째, 기저 위험 함수가 지수분포를 따르는지의 여부를 확인하기 위해 유방암 발생의 기저 위험 함수 plot을 그려 확인해보았다. 또한, 코호트 진입 연령과 공변량 간 독립의 여부를 확인하고, 코호트 진입 평균연령 등을 기술 통계량을 살펴보았다.

셋째, 폐경 전·후가 유방암 발생에 독립적인 위험 요인임을 확인하기 위해 혼란 변수들은 모형에 모두 포함하여 보정한 상태에서 다변수 분석을 시행하였다. 다변수 분석은 콕스 비례위험모형을 이용하였다.

넷째, 콕스 비례위험모형을 이용할 시, 연구 기간과 도달 연령이라는 시간 척도를 달리 준 각 네 가지 모델별의 차이를 비교 수행하였다. 결과 값은 위험비와 95% 신뢰구간(95% Confidence Interval) 값을 제시하였다.

이때, Time fixed variables를 폐경, 의료보험, 암 가족력, 체질량 지수(BMI), 알코올성 간질환, 당뇨병, 고지혈증, 고혈압으로 분석하였다.

또한, 연구 기간 방식으로 추정된 베타 계수 값 대비 도달 연령 방식으로 추정된 베타 계수 값의 변화 정도를 보기 위해 beta change percentages를 Fraction으로

계산하였다(AJJBmrm, 2001; Han, 2018).

$$Fraction = \frac{\beta_b - \beta_a}{|\beta_a|} \times 100$$

이 연구의 통계분석은 SAS version 9.4(SAS Institute Inc., Cary, NC, USA)를 사용하였고, 모든 분석의 유의수준은 5(p-value $\leq$ 0.05)로 설정하였다.

2. 연구 결과

가. 연구 대상 및 센서링 데이터 수

최종 연구 대상자 208,257명에서 추적 기간 중 유방암이 발생한 대상자는 2,541명이고, 유방암 청구내역은 없지만 사망원인이 유방암인 대상자는 3명이었다. 최종 연구 대상자에서 생존자는 195,676명이고 중도탈락 수는 10,037명으로 최종 연구 대상자에서 총 중도절단(b)는 98.8%의 비율을 보였다(표 5).

표 5. 유방암 발생 코호트 연구 대상 및 센서링 수

연구 대상 (a+b)	유방암 발생(a)			중도절단(b)		
	합계	발생자	사망자*	합계	생존자	중도탈락자†
208,257명	2,544명	2,541명	3명	205,713명	195,676명	10,037명

* 사망자 : 유방암 청구내역은 없지만, 사망원인이 유방암인 대상자를 의미
† 중도탈락자 : 사망으로 인한 중도탈락자를 의미

나. 연구 대상자의 일반적 특성

연구 대상자 총 208,257명에 대한 일반적 특성을 확인하여 표 6과 같은 결과를 얻었다. 대상은 모두 여성으로, 연구 대상자 208,257명 중 폐경 전 그룹은 64,842명(31.1%), 폐경 후 그룹은 143,415명(68.9%)이었다. 폐경 전 그룹(40-46세)의 평균 연령은 43세였으며, 출생연도는 2002년 기준 1956년도에서 1962년 범위를 보였다. 반면, 폐경 후 그룹(47-79세)의 평균 연령은 58세였으며, 출생연도는 2002년 기준 1923년에서 1955년 범위를 보였다.

의료보험은 일반 건강보험가입자(NHI)가 207,971명(99.9%), 의료급여수급자(medical aid)가 286명(0.1%)으로 두 그룹 간 유의한 차이를 보였다($p<.0017$). 암 가족력이 없는 그룹은 179,047명(86.0%), 암 가족력이 있는 그룹은 29,210명(14.0%)으로 두 그룹 간 유의한 차이를 보였다($p<.0001$). 체질량 지수(BMI)의 경우, 정상 체중(18.5-22.9 kg/㎡)군이 74,838명(35.9%)으로 가장 많이 분포했지만, 뒤이어 비만 1단계(25.0-29.9 kg/㎡) 그룹과 과체중(23.0-24.9 kg/㎡) 그룹이 각각 63,923명(30.7%), 52,347명(25.1%)을 차지하였고, 그룹 간 유의한 차이를 보일 뿐만 아니라, 경향성 검정에서도 유의하였다($p<.0001$).

2002년-2003년 사이에 알코올성 간질환이 없는 그룹은 207,025명(99.4%), 알코올성 간질환이 있는 그룹은 1,232명(0.6%)이었고 두 그룹 간 유의한 차이를 보였다($p<.0001$). 당뇨병이 없는 그룹은 187,672명(90.1%), 있는 그룹은 20,585명(9.9%)으로 두 그룹 간 유의한 차이를 보였다($p<.0001$). 고지혈증이 없는 그룹은 189,810명(91.1%), 있는 그룹은 18,447명(8.9%)으로 마찬가지로 두 그룹 간 유의한 차이를 보였다($p<.0001$). 마지막으로 고혈압이 없는 그룹은 159,313명(76.5%), 고혈압이 있는 그룹은 48,944명(23.5%)의 분포를 나타냈다. 대체적으로 유병이 없는 그룹이 더 많이 분포되어 있음을 확인하였다(표 6).

표 6. 유방암 발생 코호트 연구 대상자의 일반적 특성

Variables	All women [40-79세] (n=208,257)	Pre-menopause [40-46세] (n=64,842)	Post-menopause [47-79세] (n=143,415)	*p*-value
	n (%)	n (%)	n (%)	
Age	53.4 ± 9.8	42.9 ± 1.9	58.2 ± 8.1	
Median	52	43	58	
Range of birth(year)	1923-1962	1956-1962	1923-1955	
Medical insurance				0.0017
NHI	207,971(99.9)	64,778(99.9)	143,193(99.8)	
Medical aid	286(0.1)	64(0.1)	222(0.2)	
Cancer famliy history				<.0001
No	179,047(86.0)	53,800(83.0)	125,247(87.3)	
Yes	29,210(14.0)	11,042(17.0)	18,168(12.7)	
BMI				<.0001 *
underweight	4,682(2.3)	1,622(2.5)	3,060(2.1)	
normal	74,838(35.9)	29,959(46.2)	44,879(31.3)	
over-weight	52,347(25.1)	15,582(24.0)	36,765(25.6)	
obese class Ⅰ	63,923(30.7)	14,601(22.5)	49,322(34.4)	
obese class Ⅱ	12,467(6.0)	3,078(4.8)	9,389(6.6)	
Alcoholic Liver Disease				<.0001
No	207,025(99.4)	64,542(99.5)	142,483(99.4)	
Yes	1,232(0.6)	300(0.5)	932(0.7)	
DM				<.0001
No	187,672(90.1)	62,519(96.4)	125,153(87.3)	
Yes	20,585(9.9)	2,323(3.6)	18,262(12.7)	
Hyperlipemia				<.0001
No	189,810(91.1)	62,395(96.2)	127,415(88.8)	
Yes	18,447(8.9)	2,447(3.8)	16,000(11.2)	
Hypertension				<.0001
No	159,313(76.5)	59,983(92.5)	99,330(69.3)	
Yes	48,944(23.5)	4,859(7.5)	44,085(30.7)	

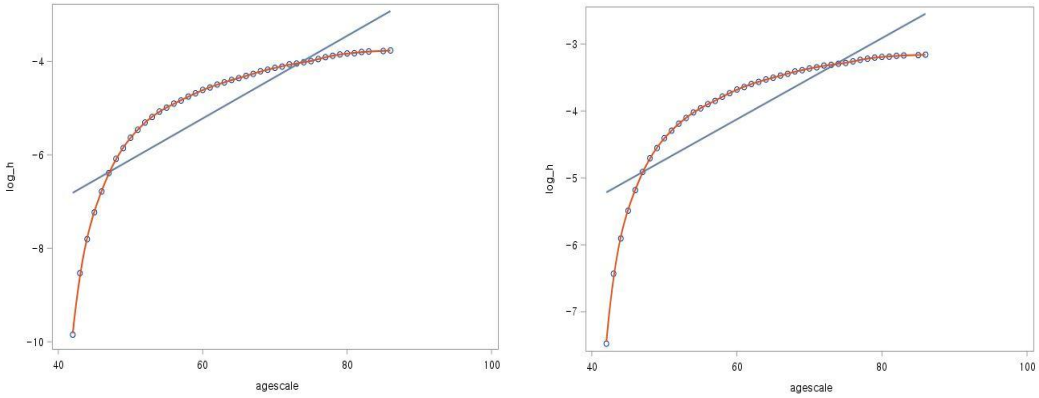
Abbreviation : BMI=body mass index; DM=diabetes mellitus.

BMI Range : underweight(<18.5 kg/㎡), normal(18.5-22.9 kg/㎡), over-weight(23.0-24.9kg/㎡), obese class Ⅰ(25.0-29.9 kg/㎡, obese class Ⅱ(≥30.0 kg/㎡)

* p_{trend}

다. 기저 위험 함수의 지수분포

코호트 진입 연령의 기저 위험 함수가 지수분포를 따르는지 확인하기 위해 <그림 9>와 같이 baseline hazard function plot을 그려 확인해보았다. [a]는 유방암 발생 시점의 나이인 도달 연령만을 고려한 것을 나타내며, [b]는 도달 연령과 연구 대상자들이 연구에 등록된 시점인 왼쪽 절단까지 적용한 그래프를 의미한다. 그래프에서 x축은 도달 연령을 y축은 $\log(h_{0a}(a))$을 의미한다.



[a] [b]

그림 9. 유방암 발생의 Baseline hazard function plot.

파란색 선 위에 직선(straight line)으로 올라가 있으면 지수분포를 따르는 것으로 볼 수 있다. [a]와 [b] 모두 파란색 선으로부터 살짝 떨어져 있는 경향으로 나타났고, [b]는 [a]와 달리, y축에 있는 시작 값과 범위가 달라지고 있음을 확인하였다.

라. 코호트 진입 연령과 공변량 간 독립성

코호트 진입 연령과 각 변수들이 독립인지, 비 독립인지 확인해보았다. 유방암 발생에 있어, 폐경(코호트 진입 평균 연령-폐경 전: 42.9세, 폐경 후: 58.2세), 암 가족력(코호트 진입 평균 연령-없음: 53.8세, 있음: 51.0세), 당뇨병(코호트 진입 평균 연령-없음: 52.8세, 있음: 59.0세), 고혈압(코호트 진입 평균 연령-없음: 51.6세, 있음: 59.4세)은 각각 모두 통계적으로 유의하였기 때문에 코호트 진입 연령과 독립적인 관계는 아님을 확인하였다.

반면, 의료보험과(p=1.000), 체질량 지수(p=0.5590), 알코올성 간질환(p=0.5239), 고지혈증(p=0.4465)에서는 각각 통계적으로 유의하지 않았으므로, 코호트 진입 연령과 독립적임을 확인하였다(표 7).

임의적으로 47세를 기준으로 폐경 전·후를 구분하였기 때문에 다른 변수들과 달리, 폐경 전의 코호트 진입 최소 연령은 40세였으며, 폐경 후는 47세였다. 이때 폐경 전의 중앙 나이는 43세이며, 폐경 후는 57세였다.

표 7. 유방암 발생 코호트 진입 연령과 공변량 간 독립성 여부

Variables	Mean SD	Median(range)	*p*-value
Menopause			<.0001
Pre-menopause	42.9±1.9	43(40, 46)	
Post-menopause	58.2±8.1	57(47, 79)	
Medical insurance			1.000 *
NHI	53.4±9.8	52(40, 79)	
Medical aid	58.5±11.4	60(40, 79)	
Cancer famliy history			0.0001
No	53.8±9.9	52(40, 79)	
Yes	51.0±8.6	49(40, 79)	
BMI			0.5590
underweight	55.4±12.1	54(40, 79)	
normal	51.9±10.1	49(40, 79)	
over-weight	53.3±9.5	52(40, 79)	
obese class Ⅰ	54.9±9.3	54(40, 79)	
obese class Ⅱ	54.5±9.5	54(40, 79)	
Alcoholic Liver Disease			0.5239
No	53.4±9.8	52(40, 79)	
Yes	53.9±9.1	53(40, 79)	
DM			0.0076
No	52.8±9.7	51(40, 79)	
Yes	59.0±9.3	60(40, 79)	
Hyperlipemia			0.4465
No	53.1±9.8	51(40, 79)	
Yes	57.0±8.8	57(40, 79)	
Hypertension			0.008
No	51.6±9.2	49(40, 79)	
Yes	59.4±9.3	60(40, 79)	

Abbreviation: DM=diabetes mellitus.

BMI Range : underweight(<18.5 kg/㎡), normal(18.5-22.9 kg/㎡), over-weight(23.0-24.9 kg/㎡), obese class Ⅰ(25.0-29.9 kg/㎡), obese class Ⅱ(≥30.0 kg/㎡)

* Fisher's exact test

마. 시간 척도 선택에 따른 4가지 모델 비교

콕스 비례위험모형을 시간 척도 선택과 연령 보정의 옵션을 달리 준 4가지 모델별로 위험비, 95% 신뢰구간, Fraction, *p*-value의 결과 값을 통해 비교하였다.

모델 1은 시간 척도를 연구 기간으로 한 모형, 모델 2는 시간 척도를 연구 기간으로 하고 연령을 보정한 모형, 모델 3은 시간 척도를 도달 연령으로 하고, 출생연도를 5년 구간으로 층화한 모형, 마지막으로 모델 4는 시간 척도를 도달 연령으로 하고 왼쪽 절단 분석을 적용한 모형으로 구성되었다.

코호트 진입 연령과 독립인 변수였던 의료보험, 체질량 지수, 알코올성 간질환, 고지혈증에서 알코올성 간질환, 고지혈증을 제외하고는 각 모델별 차이가 존재했다. 체질량 지수에서 정상(normal) 그룹은 과체중(over-weight) 그룹에 비해 출생연도를 5년 구간으로 층화한 모델 3에서는 모델 1, 2, 4와 달리 위험비(0.99)의 방향이 바뀜을 확인할 수 있었다.

주 관심 변수인 폐경에서는 모델 2에서는 다른 변수들(의료보험, 암 가족력, 체질량 지수, 알코올성 간질환, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 폐경 전 그룹은 폐경 후 그룹보다 유방암 발생 위험이 1.01배(95% 신뢰구간: 0.90-1.14) 높은 것으로 나왔고, 이는 통계적으로 유의하지 않은 결과를 보였다(p=0.817). 그러나, 모델 2의 결과와 달리 모델 1(위험비: 1.66, 95% 신뢰구간: 1.53-1.80)과 모델 3(위험비: 1.68, 95% 신뢰구간: 1.42-1.99), 모델 4(위험비: 1.55, 95% 신뢰구간: 1.35-1.77)에서는 통계적으로 유의한 결과임을 확인하였다(p<.0001).

각 모델에서 위험비의 방향이 바뀐 것은 의료보험과 당뇨였다. 모델 1에서는 다른 변수들(폐경, 암 가족력, 체질량 지수, 알코올성 간질환, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 의료급여수급자 그룹은 건강보험가입자 그룹보다 유방암 발생 위험이 0.96배(95% 신뢰구간: 0.31-2.98) 낮았고, 이는 통계적으로 유

의하지 않았다(p=0.943). 이와 달리, 모델 4에서는 다른 변수들(폐경, 암 가족력, 체질량 지수, 알코올성 간질환, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 의료급여수급자 그룹은 건강보험가입자 그룹보다 유방암 발생 위험이 1.06배(95% 신뢰구간: 0.26-2.75) 높았고, 통계적으로 유의하지 않았다(p=0.921). 당뇨의 경우, 모델 1, 3, 4와 달리 모델 2에서는 위험비의 방향이 바뀐 것을 확인할 수 있었다. 연령을 보정한 모델 2에서는 다른 변수들(폐경, 의료보험, 암 가족력, 체질량 지수, 알코올성 간질환, 고지혈증, 고혈압)을 통제했을 때, 당뇨가 있는 그룹은 당뇨가 없는 그룹에 비해 유방암 발생 위험이 1.00배(95% 신뢰구간: 0.87-1.16) 높았고, 이는 통계적으로 유의하지 않은 결과를 보였다(p=0.952).

고혈압의 경우, 다른 변수들(폐경, 의료보험, 암 가족력, 체질량 지수, 알코올성 간질환, 당뇨병, 고지혈증)을 통제했을 때, 고혈압이 있는 그룹은 없는 그룹에 비해 모델 1(95% 신뢰구간: 0.95-1.17)에서는 통계적으로 유의하지 않았지만(p=0.306), 나머지 모델 2(95% 신뢰구간: 1.07-1.31), 모델 3(95% 신뢰구간: 1.05-1.30), 모델 4(95% 신뢰구간: 1.03-1.27)에서는 통계적으로 유의한 결과를 보였다(각각, p=0.002, p=0.004, p=0.010).

추정된 베타 계수 값의 변화인 Fraction의 경우 모델별에 따라 다소 큰 차이를 보이는 변수들이 있었다. 폐경(모델 2: -97.2%, 모델 3: 2.7%, 모델 4: -13.8%)과 당뇨(모델 2: 105.0%, 모델 3: 90.6%, 모델 4: 68.6%), 그리고 고혈압(모델 2: 213.2%, 모델 3: 187.4%, 모델 4: 154.3%)으로 나타났다. 이 밖에 체질량 지수의 over-weight(모델 2: -43.6%, 모델 3: -168.2%, 모델 4: -149.6%), obese class Ⅰ(모델 2: 4.4%, 모델 3: -40.9%, 모델 4: -38.0%), obese class Ⅱ(모델 2: -26.6%, 모델 3: -81.6%, 모델 4: -65.4%) 그룹에서도 차이가 있었다.

위험비와 p-value에서 모델별에 따라 큰 차이를 보이지 않았던 변수는 암 가족력, 알코올성 간질환, 고지혈증이었다. 암 가족력의 경우 모델 4에서 다른 변수들(폐경, 의료보험, 체질량 지수, 알코올성 간질환, 당뇨, 고지혈증, 고혈압)을

통제했을 때, 암 가족력이 있는 그룹은 없는 그룹에 비해 유방암 발생 위험이 1.14배(95% 신뢰구간: 1.02-1.26) 높았고, 이는 통계적으로 유의하였다 (p=0.016). 알코올성 간질환의 경우 모델 4에서 다른 변수들(폐경, 의료보험, 암 가족력, 체질량 지수, 당뇨, 고지혈증, 고혈압)을 통제했을 때, 알코올성 간질환이 있는 그룹은 없는 그룹에 비해 유방암 발생 위험이 1.22배(95% 신뢰구간: 0.74-1.87) 높았고, 이는 통계적으로 유의하지 않았다(p=0.411). 고지혈증의 경우 모델 4에서 다른 변수들(폐경, 의료보험, 암 가족력, 체질량 지수, 알코올성 간질환, 당뇨, 고혈압)을 통제했을 때, 고지혈증이 있는 그룹은 없는 그룹에 비해 유방암 발생 위험이 1.03배(95% 신뢰구간: 0.89-1.19) 높았고, 이는 통계적으로 유의하지 않았다(p=0.662).

표 8. 유방암 발생 코호트의 4가지 모델에 따른 비교

Variable	Time scale											
	Time-on study						Attained Age					
	Not-adjusted for age(모델 1)			Adjusted for age(모델 2)			Stratified birth cohort(모델 3)			left truncation cohort entry time(모델 4)		
	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value
Menopause												
Post-menopause(ref.)	1			1			1			1		
Pre-menopause	1.66(1.53-1.80)	0.0%	<.0001	1.01(0.90-1.14)	-97.2%	0.817	1.68(1.42-1.99)	2.7%	<.0001	1.55(1.35-1.77)	-13.8%	<.0001
Medical insurance												
NHI(ref.)	1			1			1			1		
Medical aid	0.96(0.31-2.98)	0.0%	0.943	1.06(0.34-3.30)	246.4%	0.916	1.08(0.35-3.34)	280.6%	0.898	1.06(0.26-2.75)	240.7%	0.921
Cancer famliy history												
No(ref.)	1			1			1			1		
Yes	1.17(1.06-1.30)	0.0%	0.003	1.13(1.02-1.25)	-22.8%	0.023	1.13(1.02-1.26)	-21.9%	0.021	1.14(1.02-1.26)	-18.5%	0.016

Abbreviation: HR = hazard ratio; CI = confidence interval.

Variable	Time scale											
	Time-on study						Attained Age					
	Not-adjusted for age(모델 1)			Adjusted for age(모델 2)			Stratified birth cohort(모델 3)			left truncation cohort entry time(모델 4)		
	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value
BMI												
normal(ref.)	1			1			1			1		
underweight	0.85(0.63-1.14)	0.0%	0.272	0.90(0.67-1.21)	38.3%	0.498	0.91(0.68-1.23)	45.1%	0.547	0.90(0.67-1.20)	38.5%	0.499
over-weight	1.01(0.91-1.12)	0.0%	0.841	1.01(0.91-1.11)	-43.6%	0.910	0.99(0.90-1.10)	-168.2%	0.891	1.00(0.90-1.10)	-149.6%	0.921
obese class I	1.04(0.95-1.15)	0.0%	0.383	1.05(0.95-1.15)	4.4%	0.363	1.03(0.93-1.13)	-40.9%	0.607	1.03(0.93-1.13)	-38.0%	0.589
obese class II	1.03(0.87-1.23)	0.0%	0.714	1.02(0.86-1.22)	-26.6%	0.788	1.01(0.85-1.20)	-81.6%	0.946	1.01(0.85-1.20)	-65.4%	0.899
Alcoholic Liver Disease												
No(ref.)	1			1			1			1		
Yes	1.25(0.78-1.98)	0.0%	0.352	1.21(0.76-1.93)	-12.8%	0.417	1.21(0.76-1.92)	-13.9%	0.423	1.22(0.74-1.87)	-11.7%	0.411
DM												
No(ref.)	1			1			1			1		
Yes	0.94(0.81-1.08)	0.0%	0.382	1.00(0.87-1.16)	105.0%	0.965	0.99(0.86-1.15)	90.6%	0.935	0.98(0.84-1.13)	68.6%	0.784
Hyperlipemia												
No(ref.)	1			1			1			1		
Yes	1.05(0.90-1.21)	0.0%	0.545	1.05(0.91-1.21)	1.3%	0.540	1.03(0.89-1.19)	-34.6%	0.692	1.03(0.89-1.19)	-27.7%	0.662
Hypertension												
No(ref.)	1			1			1			1		
Yes	1.66(0.95-1.17)	0.0%	0.306	1.18(1.07-1.31)	213.2%	0.002	1.17(1.05-1.30)	187.4%	0.004	1.15(1.03-1.27)	154.3%	0.010

Abbreviation: HR = hazard ratio; CI = confidence interval; BMI=body mass index; DM=diabetes mellitus.

BMI Range : normal(18.5-22.9 kg/m^3), underweight(<18.5 kg/m^3), over-weight(23.0-24.9 kg/m^3), obese class I (25.0-29.9 kg/m^3), obese class II (≥30.0 kg/m^3)

Ⅳ. 알코올성 간질환이 위암 발생에 미치는 영향

1. 연구 방법

가. 연구 자료

두 번째 하위 주제에서는 검진코호트 DB를 이용하여 알코올성 간질환이 위암 발생에 미치는 영향을 파악하고 시간 척도를 연구 기간 대비 도달 연령에 따른 차이를 비교하고자 한다. 검진코호트 DB는 2002년부터 2013년까지 12년 동안 건강보험공단에 청구된 데이터를 활용했으며, 자격 DB와 진료 DB에서 명세서(20t), 상병내역(40t), 그리고 건강검진DB를 이용하였다.

남녀별에 따른 위암 유병인구 분석 결과 유병률과 발생률 모두 입원 환자에서 주상병(C16) 기준으로 위암을 정의하였을 때 통계청 인구수와 가장 비슷한 결과를 보이므로(김동욱 등, 2017), 진단 정확도를 높이기 위해 상병과 입원 여부를 함께 확인하였다. 입원 항목은 진료 DB 명세서 (T1)의과_보건기관에서 '의과 입원'과 '보건기관 입원', '정신과 입원'을 사용하였다(표 9).

표 9. 진료 DB 명세서 서식 코드별 분류

구분 변수	명세서 서식 코드별 분류		
	(T1)의과_보건기관	(T2)치과_한방	(T3)약국
서식코드 (FORM_CD)	**02 : 의과 입원** 03 : 의과 외래 **07 : 보건기관 입원** 08 : 보건기관 외래 09 : 정신과 낮병동 **10 : 정신과 입원** 11 : 정신과 외래	04 : 치과 입원 05 : 치과 외래 12 : 한방 입원 13 : 한방 기관 외	20 : 약국 조제 21 : 처방 조제

나. 연구 대상

연구 대상은 2002년 검진코호트 자격 유지자 기준 514,866명을 대상으로 추출을 시작하였다. 남자와 여자 모두에서 1차 일반 건강검진을 수검한 514,795명을 연구 대상자로 선정하였다.

일반건강검진의 특성상, 2년에 1회 주기로 수검을 받기 때문에 2002년에 건강검진을 받지 않았던 대상자는 2003년에 1차로 받은 일반 건강검진 DB를 활용했다. 반면, 매년 수검 받는 비사무직 직장가입자처럼, 2002년과 2003년 모두 수검을 받았던 대상자는 2002년에 받은 검진 결과를 기준으로 활용하였다.

연구 기간 동안 처음 진단된 환자들만 대상으로 하기 위해 2년의 워시아웃(Wash-out) 기간을 두어 2002-2003년에 사망한 대상자(1,320명)를 제외하였다. 또한, 2004년 이전 2년 동안 위암으로 청구된 이력이 있는 대상자(2,550명)는 분석에서 제외하였다. 그리고 일반 건강검진 1차 문진 항목에서 과거력 항목에 암(2,369명)을 체크한 사람도 순차적으로 제외하였다. 마지막으로 건강검진 1차 문진 항목 중 흡연 상태와 암 가족력 결측치 인원(74,056명)을 제외하고 434,500명을 최종 연구 대상자로 선정하였다. <그림 10>

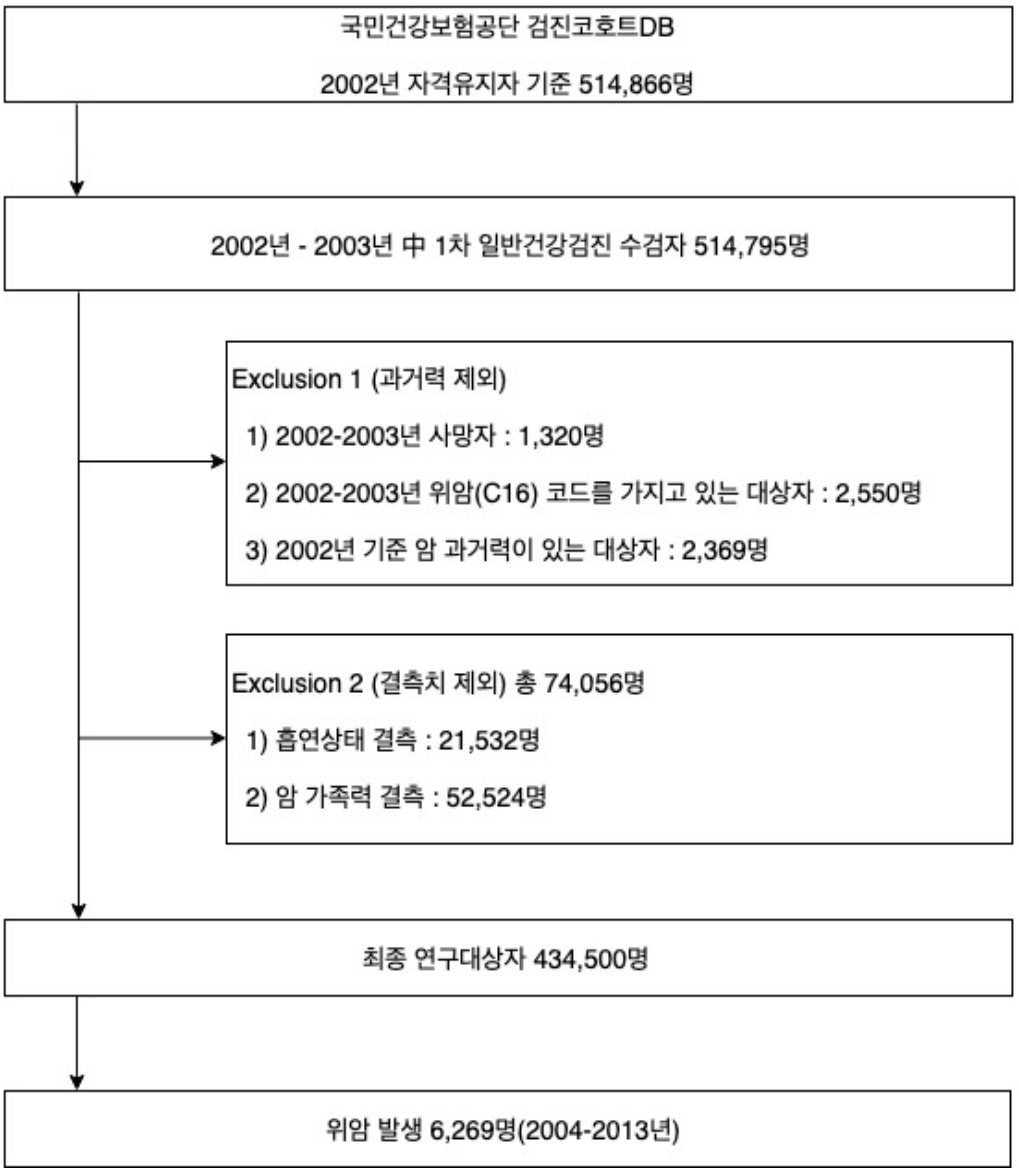


그림 10. 위암 발생 코호트의 연구 대상자 선정.

다. 변수 정의

1) 알코올성 간질환(Alcoholic Liver Disease)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 K70의 코드로 청구된 기록이 2002년-2003년 사이 주상병 또는 부상병에 존재하는 경우를 알코올성 간질환으로 정의하였다.

2) 위암(Stomach Cancer)

위암 환자는 진단 정확도를 높이기 위해 (1)주상병 또는 부상병 1순위로 위암(C16)을 진단받은 대상자 중 (2)입원환자로 정의하였다.

연구 대상자의 위암 진단은 검진코호트 진료 DB 명세서(20t)만 확인하였으며, 상병코드는 한국표준질병 사인분류(KCD) 중 C16의 코드로 2004년부터 2013년 이내에 주상병 또는 부상병 1순위로 진단받은 사람으로 사용하였다.

위암 진단받은 대상자 중 입원환자는 진료 DB 명세서 서식 코드별 분류에서 '의과 입원', '의과 외래', '보건기관 입원', '보건기관 외래', '정신과 낮 병동', '정신과 입원', '정신과 외래' 중 '의과 입원'과 '보건기관 입원', '정신과 입원'인 경우로 정의하였다.

3) 성별(Sex)

남자와 여자로 나누어 구분하였다.

4) 연령(Age)

검진코호트 중 2002년 자격 DB를 기준으로 하여 연구가 시작되는 시점에서 연구 대상자의 연령을 적용하였다. 연구 기간과 도달 연령 분석을 비교하기 위해 연령은 구간을 나누지 않고 코호트에 구축된 40세에서 79세까지의 연령

을 그대로 사용하였다.

5) 의료보험 종류(Medical Insurance Type)

자격 DB에서 지역가입자 및 직장가입자인 일반 건강보험가입자(NHI)와 의료급여수급자(Medical aid) 두 군으로 나누어 구성하였다.

6) 암 가족력(Cancer Famliy History)

일반 건강검진 1차 문진에 있는 암 가족력 변수를 활용하였으며, 가족력이 없는 군과 있는 군으로 나누어 구성하였다. 단, 2002년에 받은 건강검진을 기준으로 적용하되, 2002년에 검진을 받지 않은 대상자는 2003년에 1차로 받은 일반 건강검진 DB를 활용하고, 2002년과 2003년 모두 수검을 받았던 대상자는 2002년에 받은 건강검진 DB를 적용하였다.

7) 흡연 상태(Smoking Status)

흡연 상태는 일반 건강검진 1차 문진에서 '귀하는 담배를 어느 정도 피우십니까' 질문에 대한 응답으로, 비흡연자(Non-smoker), 과거 흡연자(Former-smoker), 현재 흡연자(Current-smoker) 3개의 그룹으로 나누어 구성하였다. 단, 2002년에 받은 건강검진을 기준으로 적용하되, 2002년에 검진을 받지 않은 대상자는 2003년에 1차로 받은 일반 건강검진 DB를 활용하고, 2002년과 2003년 모두 수검을 받았던 대상자는 2002년에 받은 건강검진 DB를 적용하였다.

8) 위궤양(Gastric Ulcer)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 K25의 코드로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 위궤양으로 정의하였다.

9) 당뇨병(Diabetes Mellitus, DM)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 제1형 당뇨를 제외한 E11, E12, E13, E14의 코드로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 당뇨병으로 정의하였다.

10) 고혈압(Hypertension)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 본태성(일차성) 고혈압(KCD-10, I10), 고혈압성 심장병(KCD-10, I11), 고혈압성 신장질환(KCD-10, I12), 고혈압성 심장 및 신장질환(KCD-10, I13), 이차성 고혈압(KCD-10, I15)으로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 고혈압으로 정의하였다.

11) 고지혈증(Hyperlipemia)

검진코호트 진료 DB의 명세서(20t)와 상병내역(40t)를 확인하였다. 한국표준질병 사인분류(KCD) 중 지질단백질 대사장애 및 기타 지질증(KCD-10, E78)으로 청구된 기록이 2002-2003년 사이 주상병 또는 부상병에 존재하는 경우를 고지혈증으로 정의하였다.

표 10. 위암 발생 코호트 변수 정의

변수		정의
독립 변수	알코올성 간질환	2002-2003년 사이 주상병 또는 부상병(K70)으로 청구된 자
혼란 변수	성별	2002년 기준 남성, 여성
	연령	2002년 기준 검진코호트 등록(Enroll)된 때 연령
	의료보험	건강보험가입자(NHI)와 의료급여수급자(medical aid)
	암 가족력	일반 건강검진 1차 결과 암 가족력 없음, 있음
	흡연 상태	일반 건강검진 1차 결과 비흡연자, 과거 흡연자, 현재 흡연자
	위궤양	2002-2003년 사이 주상병 또는 부상병(K25)으로 청구된 자
	당뇨병	2002-2003년 사이 주상병 또는 부상병(E11, E12, E13, E14)으로 청구된 자
	고혈압	2002-2003년 사이 주상병 또는 부상병(I10-I13, I15)으로 청구된 자
	고지혈증	2002-2003년 사이 주상병 또는 부상병(E78)으로 청구된 자
종속 변수	위암	연구 기간(2004-2013년)사이에 주상병 또는 부상병 1순위(C16)로 최초 입원한 자(단, 입원은 '의과 입원', '보건기관 입원', '정신과 입원' 해당)

라. Time-to-event 정의

건강보험공단 청구내역 대상자 중 연구 기간(2004-2013년) 사이에 의과 입원하여 주상병 또는 부상병으로 위암(C16)을 진료받은 청구내역이 있는 경우를 위암 발생자로 하였다. 또한, 위암 청구내역은 없지만, 사망원인이 위암인 경우도 암 발생자로 가정하였다.

Time-to-event는 검진코호트 DB의 시작일인 2002년 1월 1일을 기준으로 위암으로 청구된 첫 번째 요양 개시일 까지 시간으로 정의하였다. 위암 청구내역은 없지만, 사망원인이 위암인 대상자는 사망일을 Time-to-event로 정의했다. 2013년까지 위암으로 청구된 건이 없는 대상자는 사망 여부에 따라 아래 표 11과 같이 중도절단(Censored) 하였다.

연구 기간 대비 도달 연령을 비교 분석하기 위해 다음과 같이 도달 연령을 정의하였다. 위암 발생자는 발생한 때의 연령을, 위암이 발생하지 않은 생존자는 연구 기간이 종료되는 2013년 시 연령을 고려해주었다. 또한, 위암 청구내역은 없으나 사망원인이 위암이었던 대상자와 위암이 발생하지 않은 사망자는 사망 시 연령으로 정의하였다.

표 11. 위암 발생 코호트 Time-to-event 정의

사건발생	정의	사건이 발생할 때까지의 시간	도달 연령
위암 발생	위암 발생자	위암으로 청구된 첫 번째 요양 개시일	위암이 발생한 연령
	위암 청구내역은 없지만, 사망원인이 위암인 대상자	사망일	사망 시 연령
위암 발생하지 않음	사망자	사망일	사망 시 연령
	사망하지 않은 자(생존자)	연구 기간이 종료되는 일자 (2013년 12월 31일)	연구 기간이 종료된 2013년 시 연령

마. 분석 방법

관측 기간에 알코올성 간질환(Alcoholic Liver Disease)이 위암 발생에 영향을 미치는지 확인하고자, 위암 발생에 영향을 미칠 수 있는 가능성을 차단하기 위해 2년의 워시아웃(Wash-out) 기간을 두어 2004년 이전 2년 동안 위암 및 암 과거력을 가지고 있는 대상자는 연구에서 제외하였다.

이 연구에서는 일반적으로 생존 분석 시 흔히 사용하는 연구 기간 시간 척도를 적용한 콕스 비례위험모형과 Korn 등(1997)이 제안한 도달 연령 시간 척도를 적용한 콕스 비례위험모형을 비교함으로써 발생되는 차이를 확인하였다.

세부적인 분석은 다음과 같다. 첫째, 연구 대상자의 일반적 특성을 파악하기 위해 알코올성 간질환 유무에 따라 카이스퀘어 검정(Chi-square test)을 이용하여 빈도와 백분율을 구하고, 필요에 따라 경향성 검정을 하였다.

둘째, 기저 위험 함수가 지수분포를 따르는지의 여부를 확인하기 위해 위암 발생의 기저 위험 함수 plot을 그려 확인해보았다. 또한, 코호트 진입 연령과 공변량 간 독립의 여부를 확인하고, 코호트 진입 평균연령 등을 기술 통계량을 살펴보았다.

셋째, 알코올성 간질환이 위암 발생에 독립적인 위험 요인임을 확인하기 위해 혼란 변수들은 모형에 모두 포함하여 보정한 상태에서 다변수 분석을 시행하였다. 다변수 분석은 콕스 비례위험모형을 이용하였다.

넷째, 콕스 비례위험모형을 이용할 시, 연구 기간과 도달 연령이라는 시간 척도를 달리 준 각 네 가지 모델별의 차이를 비교 수행하였다. 결과 값은 위험비와 95% 신뢰구간(95% Confidence Interval) 값을 제시하였다.

이때, Time fixed variables를 성별, 의료보험, 암 가족력, 흡연 상태, 위궤양, 당뇨병, 고혈압, 고지혈증으로 분석하였다.

또한, 연구 기간 방식으로 추정된 베타 계수 값 대비 도달 연령 방식으로 추정된 베타 계수 값의 변화 정도를 보기 위해 beta change percentages를 Fraction으로

계산하였다(AJJBmrm, 2001; Han, 2018).

$$Fraction = \frac{\beta_b - \beta_a}{|\beta_a|} \times 100$$

이 연구의 통계분석은 SAS version 9.4(SAS Institute Inc., Cary, NC, USA)를 사용하였고, 모든 분석의 유의수준은 5(p-value $\leq$ 0.05)로 설정하였다.

2. 연구 결과

가. 연구 대상 및 센서링 데이터 수

최종 연구 대상자 434,500명에서 추적 기간 중 위암이 발생한 대상자는 6,196명이고, 위암 청구내역은 없지만 사망원인이 위암인 대상자는 73명이었다. 최종 연구 대상자에서 생존자는 399,816명이고 중도탈락 수는 28,415명으로 최종 연구 대상자에서 총 중도절단(b)는 98.6%의 비율을 보였다(표 12).

표 12. 위암 발생 코호트 연구 대상 및 센서링 수

연구 대상 (a+b)	위암 발생(a)			중도절단(b)		
	합계	발생자	사망자 *	합계	생존자	중도탈락자†
434,500명	6,269명	6,196명	73명	428,231명	399,816명	28,415명

* 사망자 : 위암 청구내역은 없지만, 사망원인이 위암인 대상자를 의미
† 중도탈락자 : 사망으로 인한 중도탈락자를 의미

나. 연구 대상의 일반적 특성

연구 대상자 총 434,500명에 대한 일반적 특성을 확인하여 표 13과 같은 결과를 얻었다. 연구 대상자 434,500명 중 알코올성 간질환이 없는 그룹(Non-ALD)이 423,812명(97.5%)이고, 있는 그룹(ALD)이 10,688명(2.5%)이다. 알코올성 간질환이 없는 그룹의 평균 연령은 50세, 있는 그룹의 평균 연령은 51세였으며, 출생연도는 모두 2002년 기준 1923년(40세)에서 1962년(79세)의 범위를 보였다.

성별은 남자가 233,040명(53.6%), 여자가 201,460명(46.4%)으로 두 그룹 간 유의한 차이를 보였다(p<.0001). 반면, 의료보험은 일반 건강보험가입자(NHI)가 434,085명(99.9%), 의료급여수급자(medical aid)가 415명(0.1%)으로 두 그룹 간 유의한 차이는 없었다(p=0.2397). 마찬가지로, 암 가족력 없는 그룹은 376,044명(86.6%), 있는 그룹은 58,456명(13.5%)으로 두 그룹 간 유의한 차이를 보이지 않았다(p=0.0701). 흡연 상태는 비흡연자가 291,951명(67.2%), 현재 흡연자 103,837명(23.9%), 과거 흡연자 38,712명(8.9%) 순으로 차지하였고, 그룹 간 유의한 차이를 보일 뿐만 아니라, 경향성 검정에서도 통계적으로 유의하였다(p<.0001).

2002년과 2003년에 위궤양이 없는 그룹은 368,798명(84.9%), 위궤양이 있는 그룹은 65,702명(15.1%)이었고, 두 그룹 간 유의한 차이를 보였다(p<.0001). 당뇨병이 없는 그룹은 391,057명(90.0%), 당뇨병이 있는 그룹은 43,443명(10.0%)으로 두 그룹 간 유의한 차이를 보였다(p<.0001). 고혈압이 없는 그룹은 341,292명(78.6%) 고혈압이 있는 그룹은 93,208명(21.5%)으로 마찬가지로 두 그룹 간 유의한 차이를 보였다(p<.0001). 마지막으로 고지혈증이 없는 그룹은 399,026명(91.8%), 고지혈증이 있는 그룹은 35,474명(8.2%)의 분포를 나타냈다(표 13).

표 13. 위암 발생 코호트 연구 대상자의 일반적 특성

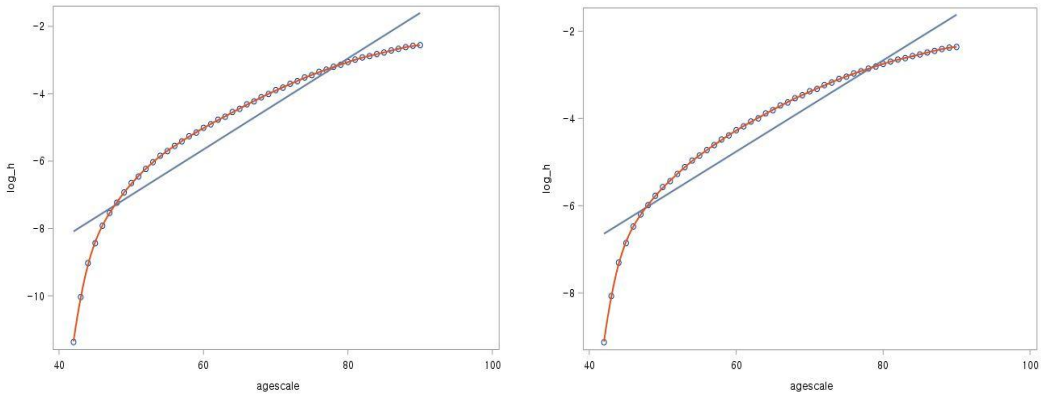
Variables	Total [40-79세] (n=434,500)	Non-ALD [40-79세] (n=423,812)	ALD [40-79세] (n=10,688)	*p*-value
	n (%)	n (%)	n (%)	
Age	52.5 ± 9.6	52.5 ± 9.6	52.7 ± 8.9	
Median	50	50	51	
Range of birth(year)		1923-1962		
Sex				<.0001
male	233,040(53.6)	223,541(52.8)	9,499(88.9)	
Female	201,460(46.4)	200,271(47.3)	1,189(11.1)	
Medical insurance				0.2397
NHI	434,085(99.9)	423,403(99.9)	10,682(99.9)	
Medical aid	415(0.1)	409(0.1)	6(0.1)	
Cancer famliy history				0.0701
No	376,044(86.6)	366,681(86.5)	9,636(87.6)	
Yes	58,456(13.5)	57,131(13.5)	1,325(12.4)	
Smoking status				<.0001 *
Non	291,951(67.2)	287,182(67.8)	4,769(44.6)	
Former	38,712(8.9)	37,234(8.8)	1,478(13.8)	
Current	103,837(23.9)	99,396(23.5)	4,441(41.6)	
Gastric ulcer				<.0001
No	368,798(84.9)	360,998(85.2)	7,800(73.0)	
Yes	65,702(15.1)	62,814(14.8)	2,888(27.0)	
DM				<.0001
No	391,057(90.0)	382,403(90.2)	8,654(81.0)	
Yes	43,443(10.0)	41,409(9.8)	2,034(19.0)	
Hypertension				<.0001
No	341,292(78.6)	333,622(78.7)	7,670(71.8)	
Yes	93,208(21.5)	90,190(21.3)	3,018(28.2)	
Hyperlipemia				<.0001
No	399,026(91.8)	390,628(92.2)	8,398(78.6)	
Yes	35,474(8.2)	33,184(7.8)	2,290(21.4)	

Abbreviation: DM=diabetes mellitus.

* p_{trend}

다. 기저 위험 함수의 지수분포

코호트 진입 연령의 기저 위험 함수(Baseline hazard function)가 지수분포(Exponential distribution)를 따르는지 확인하기 위해 <그림 11>와 같이 baseline hazard function plot을 그려 확인해보았다. [a]는 위암 발생 시점의 나이인 도달 연령만을 고려한 것을 나타내며, [b]는 도달 연령과 연구 대상자들이 연구에 등록된 시점인 왼쪽 절단까지 적용시킨 그래프를 의미한다. 그래프에서 x축은 도달 연령을 y축은 $\log(h_{0a}(a))$을 의미한다.



[a] [b]

그림 11. 위암 발생의 Baseline hazard function plot.

파란색 선 위에 직선(straight line)으로 올라가 있으면 지수분포를 따르는 것으로 볼 수 있다. [a]와 [b] 모두 점들이 파란색 선으로부터 가까이 분포되어 있는 경향을 보였다. [b]는 [a]와 달리, y축에 있는 시작 값과 범위가 달라지고 있음을 확인하였다.

라. 코호트 진입 연령과 공변량 간 독립성

코호트 진입 연령과 각 변수들 간의 독립, 비 독립인지 확인해보았다. 위암 발생에 있어, 성별(코호트 진입 평균 연령-남자: 51.8세, 여자: 53.4세), 암 가족력(코호트 진입 평균 연령-없음: 52.9세, 있음: 50.3세), 알코올성 간질환(코호트 진입 평균 연령-없음: 52.5세, 있음: 52.7세), 흡연 여부(코호트 진입 평균 연령-비흡연자: 53.3세, 과거 흡연자: 51.6세, 현재 흡연자: 50.6세), 당뇨병(코호트 진입 평균 연령-없음: 52.0세, 있음: 57.4세), 고혈압(코호트 진입 평균 연령-없음: 51.0세, 있음: 58.2세)은 각각 모두 통계적으로 유의하였기 때문에 코호트 진입 연령과 독립적인 관계는 아님을 확인하였다.

반면, 의료보험과(p=0.0631), 위궤양에서는(p=0.7913), 고지혈증(p=0.7690)은 각각 통계적으로 유의하지 않았으므로, 코호트 진입 연령과 독립적임을 확인하였다. 특히, 흡연 상태의 경우, 비흡연자-과거 흡연자-현재 흡연자 순으로 갈수록 평균 연령이 점차 낮아지고 있는 결과를 보였다(표 14).

표 14. 위암 발생 코호트 진입 연령과 공변량 간 독립성 여부

Variable	Mean SD	Median(range)	*p*-value
Sex			<.0001
male	51.8±9.3	50(40, 79)	
Female	53.4±9.8	52(40, 79)	
Medical insurance			0.0631
NHI	52.5±9.6	50(40, 79)	
Medical aid	58.6±11.3	60(40, 79)	
Cancer famliy history			0.0022
No	52.9±9.7	51(40, 79)	
Yes	50.3±8.4	48(40, 79)	
Alcoholic Liver Disease			<.0001
No	52.5±9.6	50(40, 79)	
Yes	52.7±8.9	51(40, 79)	
Smoking status			<.0001
Non	53.3±9.7	52(40, 79)	
Former	51.6±9.4	49(40, 79)	
Current	50.6±9.0	48(40, 79)	
Gastric ulcer			0.7913
No	52.2±9.5	50(40, 79)	
Yes	54.2±9.7	53(40, 79)	
DM			<.0001
No	52.0±9.4	50(40, 79)	
Yes	57.4±9.5	58(40, 79)	
Hypertension			<.0001
No	51.0±9.0	49(40, 79)	
Yes	58.2±9.5	59(40, 79)	
Hyperlipemia			0.7690
No	52.3±9.6	50(40, 79)	
Yes	55.4±9.1	55(40, 79)	

Abbreviation: DM=diabetes mellitus.

마. 시간 척도 선택에 따른 4가지 모델 비교

콕스 비례위험모형을 시간 척도 선택과 연령 보정의 옵션을 달리 준 4가지 모델별로 위험비, 95% 신뢰구간, Fraction, *p*-value의 결과 값을 통해 비교하였다.

모델 1은 시간 척도를 연구 기간으로 한 모형, 모델 2는 시간 척도를 연구 기간으로 하고 연령을 보정한 모형, 모델 3은 시간 척도를 도달 연령으로 하고, 출생연도를 5년 구간으로 층화한 모형, 마지막으로 모델 4는 시간 척도를 도달 연령으로 하고 왼쪽 절단 분석을 적용한 모형으로 구성되었다.

코호트 진입 연령과 독립인 변수였던 의료보험, 위궤양, 고지혈증 중에서 고지혈증을 제외하고는 각 모델별로 위험비 방향과 *p*-value의 차이가 존재했다. 의료보험에서는 모델 1에서 다른 변수들(성별, 암 가족력, 알코올성 간질환, 흡연 상태, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 의료급여수급자 그룹은 건강보험가입자 그룹에 비해 위암 발생 위험이 2.5배(95% 신뢰구간: 1.39-4.52) 높았고, 이는 통계적으로 유의하였다(p=0.002). 반면, 모델 2, 3, 4에서 위험비 범위는 1.65-1.68배 정도였지만, 모두 통계적으로 유의하지 않았다.

각 모델에서 위궤양, 고혈압 변수는 위험비 방향이 바뀌었다. 위궤양의 경우 모델 1(위험비: 1.04)과 달리, 모델 2(위험비: 0.95), 모델 3(위험비: 0.94), 모델 4(위험비: 0.94)에서는 정반대의 결과를 보였다. 고혈압의 경우, 모델 2, 3, 4와 달리 연령을 보정하지 않은 모델 1에서만 위험비의 방향이 다른 결과를 보였다. 모델 1에서는 다른 변수들(성별, 의료보험, 암 가족력, 흡연 상태, 알코올성 간질환, 당뇨병, 고지혈증)을 통제했을 때, 고혈압이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 1.33배(95% 신뢰구간: 1.26-1.41) 높았고, 이는 통계적으로 유의하였다(p<.0001). 이와 달리, 모델 4에서는 다른 변수들(성별, 의료보험, 암 가족력, 흡연 상태, 알코올성 간질환, 당뇨병, 고지혈증)을 통제했을 때, 고혈압이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 0.90배(95% 신뢰구

간: 0.85-0.96) 낮고, 통계적으로 유의하였다(p=0.001).

흡연 상태에서 비흡연자 대비 과거 흡연자는 모델 1에서 통계적으로 유의하지 않았던 반면, 모델 2, 3, 4에서는 유의한 결과를 보였다. 즉, 모델 1에서는 과거 흡연자는 비흡연자에 비해 위암 발생 위험이 1.08배(95% 신뢰구간: 0.99-1.17) 높았지만, 통계적으로는 유의하지 않았다(p=0.086). 그러나, 모델 1의 결과와 달리 모델 2(위험비: 1.18, 95% 신뢰구간: 1.09-1.29)과 모델 3(위험비: 1.20, 95% 신뢰구간: 1.10-1.30), 모델 4(위험비: 1.20, 95% 신뢰구간: 1.10-1.30)에서는 모두 통계적으로 유의한 결과를 보였다.

주 관심 변수인 알코올성 간질환은 모델별에 따라 큰 차이가 존재하진 않았다. 모델 4에서 다른 변수들(성별, 의료보험, 암 가족력, 흡연 상태, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 알코올성 간질환이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 1.21배(95% 신뢰구간: 1.05-1.37) 높았고, 이는 통계적으로 유의하였다(p=0.006).

특히, 당뇨의 경우, 시간 척도를 연구 기간으로 한 모델 1과 2에서는 신뢰구간 1을 포함하지 않았지만, 도달 연령으로 한 모델 3과 4는 신뢰구간 1을 포함한 결과를 보였으며, 모델 1에서 모델 4로 갈수록, 통계적으로 유의하지 않은 경향을 보였다.

추정된 베타 계수 값의 변화인 Fraction의 경우 모델별에 따라 차이가 크지 않는 경향을 보였다. 그나마, 알코올성 간질환(모델 2: 17.9%, 모델 3: 6.3%, 모델 4: 5.4%), 위궤양(모델 2: -231.8%, 모델 3: -261.3%, 모델 4: -267.9%) 변수와 과거 흡연자(모델 2: 130.9%, 모델 3: 146.7%, 모델 4: 149.0%)에서 모델 3, 4와 달리 모델 2에서 차이를 보였다.

위험비와 p-value에서 모델별에 따라 큰 차이를 보이지 않았던 변수는 성별, 암 가족력, 고지혈증이었다. 성별의 경우, 시간 척도를 도달 연령으로 하고 왼쪽 절단 분석을 적용한 모델 4에서 다른 변수들(의료보험, 암 가족력, 흡연 상태,

알코올성 간질환, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 남자는 여자에 비해 위암 발생 위험이 2.46배(95% 신뢰구간: 2.31-2.63) 높았고, 이는 통계적으로 유의하였다(p<.0001).

암 가족력의 경우 모델 4에서 다른 변수들(성별, 의료보험, 흡연 상태, 알코올성 간질환, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 암 가족력이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 1.35배(95% 신뢰구간: 1.26-1.45) 높았고, 이는 통계적으로 유의하였다(p<.0001). 고지혈증의 경우 모델 4에서 다른 변수들(성별, 의료보험, 흡연 상태, 암 가족력, 알코올성 간질환, 당뇨병, 고혈압)을 통제했을 때, 고지혈증이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 0.88배(95% 신뢰구간: 0.80-0.97) 낮았고, 이는 통계적으로 유의하였다(p<0.010).

표 15. 위암 발생 코호트의 4가지 모델에 따른 비교

Variable	Time-scale											
	Time-on-study						Attained Age					
	Not-adjusted for age(모델 1)			Adjusted for age(모델 2)			Stratified birth cohort(모델 3)			left truncation cohort entry time(모델 4)		
	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value
Sex												
Female(ref.)	1			1			1			1		
Man	2.40(2.25-2.56)	0.0	<.0001	2.49(2.34-2.66)	4.5%	<.0001	2.46(2.31-2.63)	3.0%	<.0001	2.46(2.31-2.63)	3.0%	<.0001
Medical insurance												
NHI(ref.)	1			1			1			1		
Medical aid	2.50(1.39-4.52)	0.0	0.002	1.65(0.91-2.97)	-45.8%	0.099	1.68(0.93-3.03)	-43.7%	0.087	1.65(0.86-2.84)	-45.2%	0.096
Cancer famliy history												
No(ref.)	1			1			1			1		
Yes	1.14(1.06-1.22)	0.0	0.0003	1.34(1.25-1.44)	125.7%	<.0001	1.34(1.25-1.44)	127.5%	<.0001	1.35(1.26-1.45)	129.6%	<.0001

Abbreviation: HR= hazard ratio; CI= confidence interval; DM=diabetes mellitus.

Variable	Time scale											
	Time-on study						Attained Age					
	Not-adjusted for age(모델 1)			Adjusted for age(모델 2)			Stratified birth cohort(모델 3)			left truncation cohort entry time(모델 4)		
	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value	HR(95% CI)	Fraction	*p*-value
Alcoholic Liver Disease												
No(ref.)	1			1			1			1		
Yes	1.19(1.05-1.36)	0.0	0.009	1.23(1.08-1.41)	17.9%	0.002	1.21(1.06-1.38)	6.3%	0.005	1.21(1.05-1.37)	5.4%	0.006
Smoking status												
Non(ref.)	1			1			1			1		
Former	1.08(0.99-1.17)	0.0	0.086	1.18(1.09-1.29)	130.9%	<.0001	1.20(1.10-1.30)	146.7%	<.0001	1.20(1.10-1.30)	149.0%	<.0001
Current	1.08(1.02-1.15)	0.0	0.013	1.27(1.20-1.36)	208.0%	<.0001	1.28(1.21-1.36)	216.7%	<.0001	1.29(1.21-1.37)	218.8%	<.0001
Gastric ulcer												
No(ref.)	1			1			1			1		
Yes	1.04(0.97-1.11)	0.0	0.298	0.95(0.89-1.02)	-231.8%	0.170	0.94(0.88-1.01)	-261.3%	0.093	0.94(0.88-1.01)	-267.9%	0.081
DM												
No(ref.)	1			1			1			1		
Yes	1.33(1.23-1.43)	0.0	<.0001	1.10(1.02-1.19)	-65.5%	0.011	1.08(1.00-1.17)	-72.4%	0.043	1.08(1.00-1.16)	-73.8%	0.055
Hypertension												
No(ref.)	1			1			1			1		
Yes	1.33(1.26-1.41)	0.0	<.0001	0.92(0.86-0.97)	-130.5%	0.005	0.91(0.85-0.96)	-134.8%	0.001	0.90(0.85-0.96)	-136.7%	0.001
Hyperlipemia												
No(ref.)	1			1			1			1		
Yes	0.87(0.79-0.95)	0.0	0.003	0.90(0.82-0.99)	24.2%	0.023	0.88(0.80-0.97)	13.6%	0.010	0.88(0.80-0.97)	13.8%	0.010

Abbreviation: HR= hazard ratio; CI= confidence interval; DM=diabetes mellitus.

Ⅴ. 고찰

아직 우리나라에서는 왼쪽 절단된 코호트에서 시간 척도 선택에 따른 편차를 비교하거나 밝히려는 연구는 Han(2018)의 논문 외에는 거의 찾아볼 수 없었는데, 시간 척도 선택에 따라 어떠한 잠재적 차이를 보이는지 확인해보기 위해 8가지 시나리오의 시뮬레이션을 해본 결과, 도달 연령을 시간 척도로 하고 왼쪽 절단 분석한 경우에는 실제 베타 값을 모든 시나리오에서 잘 예측하였다. 그러나, 위험 함수가 와이블 분포를 따르고, 시간 척도를 연구 기간으로 분석한 경우 공변량과 코호트 진입 연령 간 독립이면 베타 추정값은 실제 값에서 벗어난 결과를 보였다. 실제 자료인 건강보험공단 검진코호트 DB와 건강보험심사평가원의 IBD 코호트를 이용하여 비교한 결과, 연구 기간을 시간 척도로 사용했을 때 특정 상황의 경우 편차가 발생하지 않는 결과를 보이긴 했지만, 코호트 자료는 여전히 왼쪽 절단된 문제를 갖고 있기 때문에 도달 연령을 시간 척도로 사용하고 왼쪽 절단 분석하는 것을 제안했다.

국외 연구자들은 시간 척도를 도달 연령으로 사용하는 것은 일반적으로 생존분석 시 흔히 사용하는 단순 연령 보정의 효과보다 더 flexible한 통계적 방법임을 말하고 있다(Griffin BA, Anderson GL, Shih RA, et al, 2012). Thiébaut 등(2004)은 역학 코호트 연구에서 콕스 비례위험모형을 사용할 때 연령을 통제하거나, 또는 층화한다 할지라도 시간 척도를 도달 연령 대신에 연구 기간을 적용하게 되면 편차를 야기할 수 있음을 주장하였다. 연령이 통제된 경우에도 관심 공변량과 연령 간 독립성에 있어 연구 기간을 시간 척도로 사용하는 것은 편중되지 않은 추정치를 얻기 위한 충분한 조건은 아니라고 말하고 있다. 또한, 편차

는 특히 폐경 상태와 같이 시간에 따라 값이 바뀌는 변수(Time-dependent covariate)와 연령 간 상당한 연관성이 있을 때 심해질 수 있고, 게다가 연령과 공변량 간 독립적일지라도 편차를 야기할 수 있다고 지적하였다.

관심 공변량이 연령과 관련이 있는지에 따라 도달 연령보다 연구 기간을 시간 척도 사용할 때 편차가 발생할 수 있는 몇 가지 이유가 있을 수 있다. 그 이유 중 하나가 연령과 관련된 공변량의 경우 연령이 질병 발생과 무관한 경우를 제외하고, 연령은 공변량과 질병 발생 위험 사이 연관성의 혼란 변수(Confounder)로 작용할 수 있기 때문이다(Thiébaut and Bénichou., 2004).

Cheung 등(2003)은 미국 내의 암 관련 통계 정보를 제공하는 미국 국립암센터의 SEER 프로그램 데이터(NCI, 2016)를 이용하여 시간 척도로 도달 연령을 사용할 때, 유방암 진단 시 연령이 더 젊은 그룹에서 사망률과 관련이 더 높다는 것을 발견했고, 반면 연구 기간을 시간 척도로 하면 비교위험도와 신뢰구간 방향이 정반대가 나오는 결과를 확인했다.

이처럼 다양한 선행 연구에서 확인되고 있듯이 일부 연구자들은 역학 코호트 연구에서 연구 기간보다 도달 연령을 사용하는 것이 더 좋은 선택임을 주장했다(Korn et al., 1997). 이는 연령 효과(age effect)를 보정하는 간접적인 방법이라 할 수 있다(Cheung et al., 2003). 과거에는 콕스 회귀 분석의 시간 척도로서 도달 연령을 사용하는 것이 소프트웨어 패키지 한계로 인해 제한적이었지만(Cheung et al., 2003), 현재는 SAS(Allison, 1995) 또는 Strata(StataCorp, 2001)와 같은 통계 패키지에도 late entry를 고려할 수 있는 옵션이 가능하다. 이 연구에서도 왼쪽 절단된 코호트에서 연구 기간을 시간 척도로 사용한 것과 도달 연령을 시간 척도로 사용하고 왼쪽 절단 분석하는 모델 간 위험비와 95% 신뢰구간, 베타 계수 값의 변화 정도가 다른 결과를 초래할 수 있다는 것을 확인할 수 있었다.

1. 기저 위험 함수의 지수분포에 따른 차이

유방암 및 위암 발생 코호트 모두 대체적으로 지수분포를 따르는 경향을 보였지만, 위암 발생 코호트가 유방암 발생 코호트보다 좀 더 지수분포를 따르는 경향을 보였다. 실제로 위암 발생 코호트에서는 추정된 베타 계수 값의 변화인 Fraction의 경우, 전반적으로 연구 기간과 도달 연령으로 분석한 모델별에 따라 큰 차이가 없는 경향을 보였다. 위암 발생 코호트에서는 그나마 당뇨 변수에서만 모델 1에서 모델 4로 갈수록, 통계적으로 유의하지 않은 경향을 보였다. 일부 변수에서는 연령을 보정하지 않고, 연구 기간을 시간 척도로 한 모델 1에서만 차이를 보이는 일관된 결과가 나타났지만, 나머지 변수에서는 모델들은 유사한 결과를 보였다.

이처럼, 도달 연령 분석과 연구 기간 분석 방법 간 차이가 없는 경우는 Nam and Zelen(2001) 연구의 결과로 설명할 수 있다. 이 연구의 시뮬레이션 결과에 따르면 연구 기간 분석방법은 코호트 참여 전·후의 위험 함수가 같은 경우에는 비교하고자 하는 두 군 간에 코호트 참여 확률에 차이가 있어도 제1종 오류를 잘 유지하였다. 또한 코호트 참여 전·후의 위험 함수에 차이가 있어도 비교하고자 하는 두 군 간에 코호트 참여 확률에 차이가 없으면 제1종 오류가 잘 유지되었다. 이러한 상황은 생존 함수의 분포가 지수분포(Exponential distribution) 뿐만 아니라 와이블 분포(Weibull distribution)에서도 만족하였다. 즉, 도달 연령 분석과 연구 기간 분석방법 간 차이가 있는 경우는 코호트 참여 전·후 생존 함수 또는 위험 함수에 차이가 있고, 그리고 비교하고자 하는 두 군 간에 코호트 참여 확률에 차이가 있는 경우라고 생각할 수 있다. 이러한 경우는 생존 함수의 분포가 비록 지수분포를 따르더라도 연구 기간 분석방법은 제1종 오류가 잘 조정되지 않는 문제점이 있다. 연령이 증가하면서 관심 있는 사건 발생이

증가(또는 감소) 하는 만성질환 연구에서 코호트 진입 연령이 이질적이면 즉, 코호트 참여 연령군의 폭이 넓으면 코호트 참여 연령과 위험요인들 간 관련성이 존재할 가능성이 커지게 된다. 이는 비교하고자 하는 두 군 간에 코호트 참여 확률에 차이가 있을 가능성이 커지게 되는 것이고, 이에 따라 도달 연령 분석과 기존 연구 기간 분석 간에는 차이가 존재할 가능성이 커지게 된다.

2. 코호트 진입 연령과 공변량 간 독립 또는 비 독립에 따른 차이

그러나, 유방암 발생 코호트에서는 위암 발생 코호트보다 꽤 차이를 보이는 변수들이 존재하였다. 유방암 발생 코호트에서는 의료보험, 체질량 지수, 알코올성 간질환, 고지혈증이 코호트 진입 연령과 독립적이었던 변수였다. 공변량과 코호트 진입 연령 간 독립성 유무에 따라 비 독립인 경우, 독립인 경우, 독립이어도 차이를 보이는 경우인 3가지 관점에서 해석할 수 있었다.

첫째, 유방암 발생 코호트에서 공변량과 코호트 진입 연령 간 비독립적이었던 변수들은 대체적으로 모델별 차이가 있었다. 예를 들어, 폐경, 당뇨, 고혈압의 경우 위험비의 방향이 바뀌거나, 95% 신뢰구간, Fraction 값이 차이가 있었다.

둘째, 공변량과 코호트 진입 연령 간 독립이면 모델별 유사한 결과를 보였다. 예를 들어, 유방암 발생 코호트에서 알코올성 간질환과 고지혈증 변수가 해당한다.

셋째, 공변량과 코호트 진입 연령 간 독립이어도 모델에 따라 통계적 유의성의 차이가 있고, 위험비 방향이 정반대의 결과를 보이는 변수가 존재했다. 예를 들어, 유방암 발생 코호트의 체질량 지수에서는 정상(normal) 그룹은 과체중(over-weight) 그룹에 비해 다른 모델들(모델 1, 2, 4)과 달리 도달 연령으로 분석하되 출생연도를 5년 구간으로 층화한 모델 3에서만 정반대의 결과를 보이거나, 의료보험 변수는 다른 모델(모델 2, 3, 4)에 비해 모델 1에서만 위험비의 방향이 정반대였다. 이는 Thiébaut 등(2004)과 Chalise 등(2009)이 Korn 등(1997) 주장에 반박

한 연령과 공변량 간 독립적일지라도 편차를 야기할 수 있다는 연구 결과를 뒷받침한다고 볼 수 있겠다.

3. 폐경 전·후 여성에서 위험인자가 유방암 발생에 미치는 영향

기존 서양 여성을 대상으로 한 비만과 유방암 발생 및 생존율 연관성 연구들은 폐경 전 여성에서 비만이 유방암 사망 위험도를 증가시킬 수 있고, 비만군의 생존율이 비만이 아닌 군보다 낮게 나타났다(Berclaz et al., 2004; Cleveland et al., 2007). 이와 달리 한국 여성을 대상으로 한 국내 선행 연구에서는 일관적이지 않은 결과들이 많이 보고되었다. 이는 인종과 생활 및 식습관, 체형 등이 다르기 때문이라는 해석도 있었고(Min et al., 2002), 소규모 연구 대상자를 가진 병원 기반 환자 대조군 연구였기 때문에 연구 방법 및 대상 선정에서 비롯된 차이 때문이라는 해석도 존재했다(임선미 등., 2011). 임선미 등(2011)에서는 연구 시작 시점에서의 체질량 지수와 유방암 발생 위험은 고연령군에서만 경향성과 위험 증가가 관측되었고, 4년간 체질량 지수 변화에 따른 유방암 발생 위험은 통계적으로 유의한 관련성이 없었다. Lee 등(2018)은 허리둘레와 유방암 발생 관련성을 분석하였을 때, 폐경 후 여성에서는 체질량 지수를 통제하거나, 하지 않는 모델에서 허리둘레는 모두 유방암 위험과 밀접한 관련이 있었지만, 폐경 전 여성에서는 체질량 지수를 통제한 모델에서만 통계적으로 유의했음을 밝혔다. 본 연구에서 체질량 지수와 유방암 발생의 관련성을 살펴보았을 때, 체질량 지수가 정상인 군에 비해 obese class Ⅰ, Ⅱ에서는 발생 위험이 커지기는 했으나 모두 통계적으로 유의하지 않았다. 그러나, 최근 들어 유전체코호트나 국민건강보험공단 자료를 이용한 대규모 코호트에서 연구들이 보고되기 시작하였기 때문에 유방암 위험요인에 대한 유의미한 연구 결과들이 추후 이뤄질 것으로 생각된다(박수경, 2019).

그밖에 1994년부터 2007년에 구축된 서울 유방암연구(Seoul Breast Cancer Study, SeBCS)에 기반하여 한국 여성 환자-대조군 각 3,789명에 대해 연구한 Park 등(2013)은 유방암 가족력이 있는 경우 유방암 위험이 2.01배 높았고, 유방암의 분자학적 세부암 분류에 따라 50세 전후에서 약간씩 차이가 있음을 확인하였다. 또한, 2001년부터 2007년까지 서울의 3개 대학병원에 유방암 진단받은 환자와 대조군 각각 3,163명을 대상으로 연구한 Park 등(2016)에 따르면, 유방암 가족력이 있는 경우는 유방암 위험이 1.55배 높았고, 폐경 전 여성은 폐경 후 여성보다 유방암 위험이 1.74배 높았다. 본 연구에서도 모든 분석 모형에서 암 가족력 값이 유의하게 산출되어 선행연구 자료와 일치하는 것을 확인할 수 있었다. 다른 변수들(폐경, 의료보험, 체질량 지수, 알코올성 간질환, 당뇨, 고지혈증, 고혈압)을 통제했을 때, 암 가족력이 있는 그룹은 없는 그룹에 비해 유방암 발생 위험이 1.14-1.17배 높았고, 통계적으로 유의하였다.

폐경 전·후 변수에서는 모델 2를 제외하고 모델 1, 3, 4에서는 다른 변수들(의료보험, 암 가족력, 체질량 지수, 알코올성 간질환, 당뇨병, 고지혈증, 고혈압)을 통제했을 때, 폐경 전 여성은 폐경 후 여성에 비해 유방암 위험이 1.55-1.68배 높았고, 통계적으로 유의하였다(p<0.0001). 단, 모델 2에서만 통계적으로 유의하지 않은 결과(p=0.817)를 보였는데, 이는 모델 2에서 폐경 전·후 변수를 보정된 연령 즉, 코호트 등록 당시 연령으로 그룹화를 해준 것이 값에 영향을 준 것 같다. 또한, 동반질환에서 유의한 결과를 보인 것은 고혈압 변수였다. Heo 등(2019)은 2010년부터 2015년까지 6년 동안 건강보험심사평가원에 등록된 유방암 생존자 89,953명을 대상으로 대사성 질환에 대하여 확인한 결과, 유방암 환자의 약 30%에서 당뇨병, 고혈압, 고지혈증 등 만성질환을 경험하는 것으로 나타났다. 이는 유방암 생존자에서 만성질환과의 연관성과 관리의 필요성을 시사한다.

4. 알코올성 간질환 유무에 따라 위암 발생에 미치는 영향

알코올성 간질환 유무에 따라 위암 발생에 미치는 영향을 밝히는 국내 논문은 거의 찾기 어려워 직접적인 비교는 할 수 없었다. 대신 우리나라에서 알코올 섭취와 위암 발생과의 관계는 다수의 논문에서 연관이 있는 것으로 확인되었다. Sung 등(2007)은 1996-2002년 국민건강보험공단 코호트의 한국인 남성 669,570명을 분석한 결과, 알코올을 일일 25g 이상 섭취하는 경우 비음주자에 비해 위암과의 연관성이 1.2배 유의하게 증가하였다. 이때 분석에서는 코호트 진입 시 연령을 보정해주었다. 한국인 다기관 암 코호트 연구인 Jung 등(2012)은 20세 이상 16,320명의 대상자에서 알코올 소비 습관과 사망률 사이의 연관성을 분석한 결과, 알코올 섭취자는 술을 마시지 않은 사람에 비해 전체 사망률의 위험이 1.72배로 증가했다. 또한, 알코올을 주당 504.01g 초과하여 섭취하는 경우는 주당 0.01-9.9g 섭취하는 경우에 비해서 위암과의 연관성이 2.93배 증가하였으며 유의한 결과를 보였다. 이때 연령 보정은 10세 구간으로 범주화시킴으로써 분석하였다. Choi 등(2017)에서는 국민건강보험공단의 2년마다 검진받는 검진 코호트를 이용하여 23,323,730명을 대상으로 음주와 식도, 위암, 대장암 위험 사이의 연관성을 조사하였다. 다른 독립변수와 함께 연령을 보정한 결과, 비음주자에 비해 경도의 음주자의 경우 위암의 위험이 1.05배 증가했고, 과도 음주자의 경우 1.24배 증가하였다. 이처럼 다양한 선행 연구에서 확인되고 있듯이 음주는 위암 발생에 영향을 미치는 중요한 변수이다. 본 연구에서는 음주 여부에 대한 변수를 모형에 넣진 않았지만, 선행연구에 따라 알코올 섭취에 영향을 받는 알코올성 간질환 변수를 모형에 넣어 분석했을 때 모든 분석 모형에서 유의하게 산출되었다. 알코올성 간질환의 유무가 위암 발생 연관성을 나타내는 위험비, 95% 신뢰구간 값이 모델별에 따라 큰 차이가 존재하진 않았고, 다른 변수들(성별, 의료보험, 암 가족력, 흡연 상태, 당뇨병, 고

지혈증, 고혈압)을 통제했을 때 알코올성 간질환이 있는 그룹은 없는 그룹에 비해 위암 발생 위험이 1.19-1.23배(모델 1, 2, 3, 4) 높았고, 모두 통계적으로 유의한 결과를 보였다.

5. 연구의 제한점

이 연구를 진행하면서 연구 방법과 관련한 몇 가지 제한점이 존재한다. 첫째, 유방암 및 위암 발생 코호트 진입 연령 범위가 40-79세였기 때문에 진입 연령이 이질적인지, 동질적인지 확인하기에 어려움이 있었으므로 이에 따라 비교하고자 하는 두 군 간에 코호트 참여 확률 차이를 파악하는 데 한계가 있다.

둘째, 연구 자료가 갖고 있는 특성상 유방암과 위암에 영향을 줄 수 있는 임상자료를 활용하지 못했다. 특히, 유방암의 경우 많은 인자가 질병에 관여하는 다요인 복합 질병이지만, 기존에 확립된 유방암의 위험요인들은 주로 여성 호르몬 노출과 관련성이 있기 때문에 빠른 초경, 늦은 폐경, 출산 여부 등의 고려가 필요한데(박수경, 2019), 이에 대한 정보를 얻을 수 없어 유방암 발생에 대한 위험요인의 영향을 임상적인 측면을 고려하여 평가할 수 없었다.

셋째, 건강보험공단의 건강검진 설문지에는 폐경 여부를 확인할 수 있는 조사 문항이 부재하였으므로 폐경 여부에 대한 정확한 정보를 반영하지 못한 연구의 제한점이 존재한다. 대안으로 2004년 한국 여성의 건강통계 결과 보고를 인용하여 폐경 전·후를 구분하였으나 개인에 대한 정확한 폐경 정보를 적용하지 않았으므로 임선미 등(2011)에서 지적했던 것처럼, 오분류 편의(Misclassification bias)가 발생했을 수 있으며, 이는 연구 결과에 영향을 줄 수 있다. 실제로 다른 변수들을 통제했을 때, 연령을 보정하고 연구 기간으로 분석한 모델 2에서만 p-value가 0.817로 통계적으로 유의하지 않았고, 모델 1, 3, 4에서는 모두 통계적으로 유의했다($p<.0001$). 하지만, 이런 오분류 편의는 실제 폐경 전 그룹과 후 그룹에서

비 편향적(non-differential)으로 발생했을 것으로 생각되므로 연구 결과가 한 방향으로 비뚤어지지는 않았을 것으로 생각된다(임선미 등, 2011). 그렇지만, 폐경 전·후에 따른 유방암 발생에 대한 위험요인의 영향을 정확히 파악하기 위해 건강검진 설문지에 폐경 여부를 묻는 문항이 생성되길 제안해본다. 또한 경구피임약 등의 복용에 대한 정보가 포함되었다면 유방암 예방 및 발생에 어떤 영향을 미쳤는지 추정해볼 수 있었겠지만 이를 반영하지 못했다.

넷째, 시간에 따라 변수의 값이 바뀌는 변수(Time-dependent covariate)를 고려하여 분석하지 못했기 때문에 체질량 지수 변화와의 유방암 관련성, 흡연 상태에 따른 위암 발생의 관계를 파악하고 평가할 수 없었다.

이러한 한계에도 불구하고 이 연구는 기존에 국내에서는 연구가 많이 이루어지지 않은 시간 척도 선택에 따른 차이를 비교했다는 점에서 의의가 있으며, 특히 역학 코호트 연구에서 잠재적 편차를 해결하기 위한 방법의 하나인 도달 연령 방법과 왼쪽 절단된 형태를 고려하여 분석했다는 점에서 의미가 있다 할 수 있겠다.

하지만 질병과 위험 요인의 연관성을 가장 잘 설명할 수 있는 적절한 시간 척도를 찾기 위한 연구가 많이 필요한 부분이라고 할 수 있다. 국외 선행연구자들은 비례위험모형에서 시간 척도를 선택하고 적절성을 평가하는 것은 간단하지 않다고 보고하고 있기 때문이다(Chalise et al., 2009). 특히 연령은 보건학에서 만성질환 발생 위험에 영향을 주는 대표적인 혼란 변수로 연령 효과의 통제는 코호트 연구에서 중요한 문제였다. 이때 관심 있는 독립변수 Z와 코호트가 등록되는 시점인 ω와 서로 관련이 있다면, 기간 차이 바이어스(Length bias)가 존재하게 된다. 보통 임상 연구에서 clinical event가 생기면, 이 시점에서 왼쪽 절단이 발생했다고 본다. 임상적으로 중요한 사건을 치료(Treatment)라고 본다면, 예를 들어, HIV 환자에서 치료가 시작되면 그 시점을 기준으로 환자의 생존 분포(Survival distribution)가 바뀌면서 사건을 경험한 사람과 경험하지 않는 사람

들 간의 비교하는 과정에서 기간 차이 바이어스가 생긴다. 그래서 생존 분포의 패턴이 바뀌는 시점에서 절단(Truncation)시켜서 분석한다. 하지만, 관찰 코호트 연구는 clinical event가 생기는 임상 연구는 아니지만, 연구에 등록되는 시점까지는 살아있어야 코호트에 포함되기 때문에 생존 분포가 바뀔 수 있게 되는 것이다. 시간 척도를 연구 기간으로 분석할 때는 코호트 진입 연령을 보정함으로써 기간 차이 바이어스를 제거할 수는 있지만, 대상자들은 각기 다른 시점과 연령에서 연구에 등록되기 때문에 코호트가 왼쪽 절단됨으로써 발생하는 편차 문제를 해결하지는 못한다(Han, 2018). 따라서, 콕스 비례위험모형을 사용하는 코호트 연구에서는 도달 연령을 시간척도로 하여 왼쪽 절단 분석하는 것을 제안한다.

Ⅵ. 결 론

이 연구는 두 가지 하위 주제에 대해 시간 척도 연구 기간 분석방법과 도달 연령 분석방법을 달리 선택함에 따라 발생하는 콕스 비례위험모형 간 위험비와 베타 계수 값의 차이를 비교하고 분석하였다.

기저 위험 함수가 지수 분포를 따르는 경향을 보이거나 또는 코호트 진입 연령과 공변량 간 독립이면 두 시간 척도 분석 결과의 편차는 줄어드는 결과를 보였다. 그러나, 코호트 진입 연령과 공변량 간 독립일지라도 간혹 차이를 보이는 변수들이 존재했다. 눈여겨보아야 할 점은 시간 척도를 달리 선택함에 따라 위험비의 방향이 바뀌거나 통계적 유의성이 달라지는 차이가 존재한다는 것인데 이는 질병과 위험 요인의 연관성을 규명해내는 역학 코호트 연구 에서 중요한 이슈가 될 수 있다.

따라서, 왼쪽 절단된 코호트에서 연구 기간을 시간 척도로 사용하면 왼쪽 절단됨으로써 발생하는 편차 문제가 제기되므로 코호트 연구에서 콕스 비례위험모형을 사용하는 경우 도달 연령을 시간 척도로 하여 왼쪽 절단 분석하는 것을 제안한다. 또한, 여러 위험요인은 연령의 증가에 따라 누적되어 그 영향성이 증가할 가능성이 있기 때문에 질병과 위험 요인의 연관성을 가장 잘 설명할 수 있는 적절한 시간 척도를 찾기 위한 많은 연구가 이루어져야 할 것이다.

참고문헌

고광필. 우리나라 위암의 역학. 대한의사협회지 2019;62(8):398-406.

국민건강보험공단. 건강검진코호트DB 사용자 매뉴얼, 2016.

김동욱, 이선미, 임현선, 최정규, 박해용, 육태미, 강민진, 홍정화, 한규태, 배세진. 건강보험 청구자료에 근거한 질병의 조작적 정의에 관한 연구. 백석기획, 2017.

박수경. 한국인에서 유방암의 역학적 특성. 대한의사협회지 2019;62(8):424-436.

송민교, 이휘원, 강대희. 한국인에서의 위암의 역학적 특성과 위암검진. 대한의사협회지 2015;58(3):183-190.

오현경. 비만 및 기타 유방암 위험인자와 유방암 생존율과의 관련성 연구[석사학위 논문]. 성균대학교 일반대학원; 2010.

임선미, 허남욱, 김현창, 강대용, 서일. 체질량 지수와 유방암발생의 관련성. 한국보건정보통계학회지 2011;36(1):39-49.

정한영, 이숙향. 혈당강하제 단독요법 투여 당뇨병환자에서 암발생률 평가: 후향적 코호트 연구. 한국임상약학회지 2019;29(3):186-92.

한국유방암학회. 2018 유방암백서, 2018.

한국중앙암등록본부. 국가암등록사업 연례 보고서(2016년 암등록통계), 보건복지부, 2018.

Allison PD. Survival analysis using the SAS system. Cary, NC: SAS Institute Inc. 1995.

Berclaz G, Li S, Price KN, Coates AS, Castiglione-Gertsch M, Rudenstam CM, et al. Body mass index as a prognostic feature in operable breast cancer: the International Breast Cancer Study Group experience. Annals of Oncology 2004;15(6):875-84.

Berg VD, Gerard J. Inference for shared-frailty survival models with left-truncated data: discussion paper series. Econometric Reviews 2011;35(6).

Bilker WB, Wang MC. A semiparametric extension of the Mann - hitney test for randomly truncated data. Biometrics 1996; 52(1):10 - 20.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394-424.

Breslow, N. Covariance analysis of censored survival data. Biometrics. 1974;30(1):89-99.

Chalise P, Chicken E, McGee D. Time Scales in Epidemiological Analysis. An Empirical Comparison. 2009;00:1-13.

Cheung YB, Gao F, Khoo KSJJoce. Age at diagnosis and the choice of survival analysis methods in cancer epidemiology. 2003;56(1):38-43.

Chi, Y., Tsai, W. Y. and Chiang, C. L. Testing the equality of two survival functions with right truncated data, Statistics in Medicine. 2007;26(4):812-27.

Choi YJ, Lee DH, Han KD, Kim HS, Yoon H, Shin CM, Park YS, Kim N. The relationship between drinking alcohol and esophageal, gastric or colorectal cancer: a nationwide population-based cohort study of South Korea. PLoS One 2017;12(10):e0185778.

Cleveland RJ, Eng SM, Abrahamson PE, Britton JA, Teitelbaum SL, Neugut AI, et al. Weighy gain prior to diagnosis and survival from breast cancer. Cancer Epidemiol Biomarkers Prev 2007;16(9):1803-11.

Commenges D, Letenneur L, Joly PJAjoe. Re:"Serum transferrin saturation, stroke incidence, and mortality in women and men. The NHANES I Epidemiologic Followup Study". 1997;146(8):683-84.

Correa P. Human gastric carcinogenesis: a multistep and multifactorial process First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res 1992;52(24):6735 40.

Cox DR. Regression models and life-tables. Breakthroughs in statistics: Springer 1992:527-41.

Ferro A, Morais S, Rota M, Pelucchi C, Bertuccio P, Bonzi R, Galeone C, Zhang ZF, Matsuo K, Ito H, Hu J, Johnson KC, Yu GP, Palli D, Ferraroni M, Muscat J, Malekzadeh R, Ye W, Song H, Zaridze D, Maximovitch D, Aragones N, Castano-Vinyals G, Vioque J, Navarrete-Munoz EM, Pakseresht M, Pourfarzi F, Wolk A, Orsini N, Bellavia A, Hakansson N, Mu L, Pastorino R, Kurtz RC, Derakhshan MH, Lagiou A, Lagiou P, Boffetta P, Boccia S, Negri E, La Vecchia C, Peleteiro B, Lunet N. Tobacco smoking and gastric cancer: meta-analyses of published data versus pooled analyses of individual participant data (StoP Project). Eur J Cancer Prev 2018;27(3):197-204.

Fieberg J, DelGiudice GD Estimating age-specific hazards from wildlife telemetry data. Environmental and Ecological Statistics. 2009;18(2):209 - 22.

Gilbert SL, Lindberg MS, Hundertmark KJ, Person DK, Dead before detection: addressing the effects of left truncation on survival estimation and ecological inference for neonates. 2014;5(10):992-1001.

GLOBOCAN 2012. Estimated cancer incidence, mortality and prevalence worldwide in 2012. Lyon: International Agency for Research on Cancer. Accessed September 1st, 2016. Available from http://globocan.iarc.fr/Pages/fact_sheets_cancer.

Griffin BA, Anderson GL, Shih RA, et al. Use of alternative time scales in Cox proportional hazard models: implications for time varying environmental exposures. 2012;31(27):3320-27.

Guggenheim DE, Shah MA. Gastric cancer epidemiology and risk factors. J Surg Oncol 2013;107(3):230-6.

Han MK. Choice of time scale for proportional hazard models in cohort data: Simulation and real data analyses[dissertation]. Graduate school of public health yonsei university; 2018.

Heo JS, Chun MS, Oh YT, Noh OK, Kim LY. Metabolic comorbidities and medical institution utilization among breast cancer survivors: a national population-based study. Korean J Intern Med 2019.

In JY and Lee DK, Survival analysis: Part I - analysis of time-to-event. Korean Journal of Anesthesiology 2018: 187.

Jee SH, Ohrr H, Sull JW, Yun JE, Ji M, Samet JM. Fasting serum glucose level and cancer risk in korean men and women. JAMA 2005;293(2):194-202.

Jeong SH, An YS, Choi JY, Park B, Kang D, Lee MH, Han W, Noh DY, Yoo KY, Park SK. Risk reduction of breast cancer by childbirth, breastfeeding, and their interaction in Korean women: heterogeneous effects across menopausal status, hor-mone receptor status, and pathological subtypes. J Prev Med Public Health 2017;50(6):401-10.

Jiang Y. Estimation of Hazard Function for Right Truncated Data. Georgia State University, Master Thesis 2011.

Jung EJ, Shin A, Park SK, Ma SH, Cho IS, Park B, Lee EH, Chang SH, Shin HR, Kang D, Yoo KY. Alcohol consumption and mortality in the Korean Multi-Center Cancer Cohort Study. J Prev Med Public Health 2012;45(5):301-8.

Kang SY, Kim YS, Kim Z, Kim HY, Lee SK, Jung KW, Youn HJ; Korean Breast Cancer Society. Basic Findings Regarding Breast Cancer in Korea in 2015: Data from a Breast Cancer Registry. J Breast Cancer. 2018;21(1):1-10.

Klein, JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer, 2003.

Kom EL, Graubard BI, Midthune DJAjoe. Time-to-event analysis of longitudinal followup of a survey: choice of the time-scale. 1997;145(1):72-80.

Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, Lunet N. Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. Cancer Causes Control 2008;19(7):689-701.

Lamarca R, Alonso J, Gomez G, et al. Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. 1998;53(5):337-43.

Lee KR, Hwang IC, Han KD, Jung J, Seo MH. Waist circum-ference and risk of breast cancer in Korean women: A nation-wide cohort study. Int J Cancer 2018;142(2):1554-9.

Min YK, Park CM, Kim WB, Cho SJ, Kim AR, Kim NR et al. Distribution and Prognostic Effect on Adjuvant Hormone Therapy of Body Mass Index (BMI) in Korean Breast Cancer Patients. J Korean Surg Soc 2002; 62(4): 275-281 (Korean).

Minami Y, Kanemura S, Oikawa T, Suzuki S, Hasegawa Y, Miura K, Nishino Y, Kakugawa Y, Fujiya T. Associations of cigarette smoking and alcohol drinking with stomach cancer survival: A prospective patient cohort study in Japan. Int J Cancer 2018;143(5):1072-85.

Ministry of Health and Welfare. Korea Women's Health Statistics, 2004.

Nam CM, Zelen MV. Comparing the survival of two groups with an intermediate clinical event. 2001;7(1):5-19.

NCI. Surveillance, Epidemiology, and End Results (SEER) Program, www.seer.cancer.org. 2016.

Park B, Choi JY, Sung HK, Ahn C, Hwang Y, Jang J, Lee J, Kim H, Shin HR, Park S, Han W, Noh DY, Yoo KY, Kang D, Park SK. Attribution to heterogeneous risk factors for breast cancer subtypes based on hormone receptor and human epidermal growth factor 2 receptor expression in Korea. Medicine (Balti-more) 2016;95(4):e3063.

Park B, Ma SH, Shin A, Chang MC, Choi JY, Kim SW, Han W, Noh DY, Ahn SH, Kang DH, Yoo KY, Park SK. Korean risk assessment model for breast cancer risk prediction. PLoS One 2013;8(10):e76736.

Park B, Park S, Shin HR, Shin A, Yeo Y, Choi JY, Jung KW, Kim BG, Kim YM, Noh DY, Ahn SH, Km JW, Kang S, Kim JH, Kim TJ, Kang D, Yoo KY, Park SK. Population attributable risks of modifiable reproductive factors for breast and ovarian cancers in Korea. BMC Cancer 2016;16(5).

Park S, Shin HR, Lee B, Shin A, Jung KW, Lee DH, Jee SH, Cho SI, Park SK, Boniol M, Boffetta P, Weiderpass E. Attributable fraction of alcohol consumption on cancer using population-based nationwide cancer incidence and mortality data in the Republic of Korea. BMC Cancer

2014;14:420.

Pencina M, Larson M, and D'Agostino R. Choice of time scale and its effect on significance of predictors in longitudinal studies. Statistics in Medicine 2007;26(6):1343-59.

Shin CM, Kim N, Yang HJ, Cho SI, Lee HS, Kim JS, Jung HC, Song IS. Stomach cancer risk in gastric cancer relatives: interaction between Helicobacter pylori infection and family history of gastric cancer for the risk of stomach cancer. J Clin Gastroenterol 2010;44(2):e34-9.

StataCorp. Stata reference manual. Release 6. College Station, TX: Stata Press; 2001.

Sung NY, Choi KS, Park EC, Park K, Lee SY, Lee AK, Choi IJ, Jung KW, Won YJ, Shin HR. Smoking, alcohol and gastric cancer risk in Korean men: the National Health Insurance Corporation Study. Br J Cancer 2007;97(5):700-4.

Thiébaut AC, Bénichou JJSim. Choice of time scale in Cox's model analysis of epidemiologic cohort data: a simulation study. 2004;23(24):3803-20.

Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends-An Update. Cancer Epidemiol Biomarkers Prev

2016;25(1):16-27.

Tramacere I, Negri E, Pelucchi C, Bagnardi V, Rota M, Scotti L, Islami F, Corrao G, La Vecchia C, Boffetta P. A meta-analysis on alcohol drinking and gastric cancer risk. Ann Oncol 2012;23(1):28-36.

Tsugane S, Sasazuki S. Diet and the risk of gastric cancer: review of epidemiological evidence. Gastric Cancer 2007;10(2):75-83.

Using discrete-time event history fertility models to simulate total fertility rates and other fertility measures, Population and Policy Review. J Bus Ventur 2012;32(4):477-92.

Van Hook, J. and Altman, C.E. Using discrete-time event history fertility models to simulate total fertility rates and other fertility measures, Population and Policy Review. 2013;32(4):585-610.

Vickers AJJBmrm. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. 2001;1(1):6.

Vigneri P, Frasca F, Sciacca L, Pandini G, Vigneri R. Diabetes and cancer. Endocr Relat Cancer 2009;16(4):1103-23.

Wang MC. Nonparametric estimation from cross-sectional survival data.

Journal of the American Statistical Association 1991;86(413):130 - 43.

WHO Expert Consultation. Appropriate bodymass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363(9403):157-63.

World Cancer Report 2014 : International Agency for Research on Cancer. Lyon: 2014.

World Cancer Research Fund, American Institute for Cancer Research. Diet, nutrition, physical activity and stomach cancer [Internet]. London: World Cancer Research Fund International; 2018 [cited 2019 Jul 12]; Available from: https://www.wcrf.org/sites/default/files/Stomach-cancer-report.pdf.

Yaghoobi M, Bijarchi R, Narod SA. Family history and the risk of gastric cancer. Br J Cancer 2010;102(2):237-42.

Yang T, Aldrich HE. Out of sight but not out ofmind: Why failure to account for left truncation biases research on failure rates. Journal of Business Venturing, 2012;27(4):477 - 92.

Yi SW, Sull JW, Linton JA, Nam CM, Ohrr H. Alcohol consumption and digestive cancer mortality in Koreans: the Kangwha Cohort Study. J Epidemiol 2010;20(3):204-11.

Yoo KY, Kang D, Park SK, Kim SU, Shin A, Yoon H, Ahn SH, Noh DY, Choe KJ. Epidemiology of Breast Cancer in Korea: Occurrence, High-Risk Groups, and Prevention. J Korean Med Sci 2002;17(1):1-6.

Zanghieri G, Di Gregorio C, Sacchetti C, et al. Familial occurrence of gastric cancer in the 2 year experience of a population based registry. Cancer 1990;66(9):2047 51.

= Abstract =

Comparison of Differences According to Time Scale Selection in Left-truncated Cohorts

Bi Dan
Department of Biostatistics
Graduate School of Public Health
Yonsei University

(Directed by Professor Chung Mo Nam, Ph.D.)

Research Background

Time scale selection is an important issue in survival analysis in epidemiological cohort studies. Survival and observational studies using Cox's proportional hazard model have primarily used time-on study time scales, which were enrolled in the study and calculated the time to event. Age is a representative confounding variable that affects the risk of disease in health

science. However, the use of time-on studies raises the question of whether all subjects registered at the same time in the study are given the same follow-up period and therefore the risks are all the same. In addition, foreign prior researchers found that the subjects are registered in the study at different time points and ages, resulting in left truncation, which increases survival estimates and suggests the possibility of biasing the estimated effects of covariates. This study was conducted to compare whether there is any bias in selecting different time scales by applying the attained age time scale that calculates the time from birth of age 0 to the occurrence of an event.

Subject and Method

A time scale model was compared using the breast cancer cohort and the gastric cancer occurrence cohort. The breast cancer occurrence cohort was the subject of risk factors on the development of breast cancer in pre and postmenopausal women. The gastric cancer occurrence cohort was related to the effects of alcoholic liver disease on the development of gastric cancer. Both the study subjects were adults over the age of 40 from 2002 to 2013 using the National Health Insurance Corporation examination cohort DB. Whether the baseline hazard function follows an exponential distribution and the independence between cohort entry age and covariates were identified. The Cox proportional hazard model was used to compare the differences by calculating the hazard ratios, 95% confidence intervals, and the Fractions representing beta change percentages. A total of four models were used for the analysis. Model 1 is a time-on study with

time scale, model 2 is a time-on study with time scale with age adjusted and model 3 is that with time scale as attained age and composed of stratified model of birth year with 5 year intervals. Finally, model 4 consisted of model with left truncation applied and time scale as attained age.

Research Result

First, in the left truncation cohort, the risk ratio direction is changed, or the 95% confidence interval and Fraction value are changed between the models using the time-on study as the time scale and the attained age as the time scale, and the left truncation analysis.

Second, even if the baseline hazard function follows an exponential distribution or the cohort entry age and the covariate are independent, the bias of the two time scale analysis results is reduced. However, there were some differences in both breast and gastric cancer occurrence cohorts, even if independent of cohort entry age and covariates.

Third, in the menopausal variables of breast cancer occurrence cohort, except for model 2, in models 1, 3, and 4, when other variables (medical insurance, cancer family history, body mass index, alcoholic liver disease, diabetes mellitus, hyperlipidemia, hypertension) were controlled, premenopausal women were 1.55-1.68 times more likely to have breast cancer than postmenopausal women and were statistically significant.

Fourth, there was no significant difference between the gastric cancer occurrence cohort in the presence of alcoholic liver disease and the risk

ratio and 95% CI values, which were associated with gastric cancer. When controlling for other variables (gender, medical insurance, cancer family history, smoking status, diabetes mellitus, hyperlipidemia, hypertension), those with alcoholic liver disease were 1.19-1.23 times higher (model 1, 2, 3, 4) than those without alcoholic liver disease, all with statistically significant results.

Conclusion

The difference in the direction of hazard ratios and the statistical significance in selecting different time scales can be an important issue in epidemiological cohort studies that identify the association between disease and risk factors.

Using the time-on study as a time scale in a left truncation cohort raises the problem of bias caused by left truncation. Therefore, when using the Cox proportional hazard model in a cohort study, we propose a left truncation analysis using the attained age as a time scale and subsequent studies should be made to select the appropriate time scale.

Key Words: left truncation, time scale, time-on study, attained age, cohort, age-adjustment, Cox's proportional hazard model