

연구보고서 2017-20-027

www.nhimc.or.kr

# 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발(4차년도)

조경희 · 유보라 · 김동욱 · 한규태 · 이재우 · 조현순 · 신미연  
김 열 · 박보영 · 배철영 · 강영곤 · 김대중 · 하경화 · 서용성

# NHIS

2017 NHIS Ilasan Hospital  
National Health Insurance Service Ilasan Hospital



국민건강보험 일산병원 연구소

연구보고서

2017-20-027

# 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발(4차년도)

조경희·유보라·김동욱·한규태·이재우  
조현순·신미연·김열·박보영  
배철영·강영곤·김대중·하경화·서용성

[저 자]

책임 연구자:	국민건강보험 일산병원 가정의학과	조경희
공동 연구원:	국민건강보험 일산병원 연구분석팀	유보라
	국민건강보험 일산병원 연구분석팀	김동욱
	국민건강보험 일산병원 연구분석팀	한규태
	국립암센터	조현순
	국립암센터	김열
	국립암센터	박보영
	국립암센터	신미연
	아주대학교	김대중
	아주대학교	하경화
	명지병원	서용성
	충북대학교	이재우
	메디에이지	배철영
	메디에이지	강영곤
보조 연구원:	일산병원	곽정은

본 연구보고서에 실린 내용은 국민건강보험 일산병원의 공식적인 견해와 다를 수 있음을 밝혀둡니다.

## 머리말

의료 이용자들의 건강에 대한 관심은 점차 증가하고 있으며 자신의 건강과 관련된 요인들을 스스로 결정하고자 하는 욕구가 증가하고 있다. 생활수준이 향상되고 의학기술이 급속하게 발달함에 따라 질병의 예방중심의 의료로의 보건으로 패러다임이 전환되고 있다. 국민들에게 맞춤형의 올바른 건강 정보를 제공하기 위하여 전국민 건강검진 관련 정보를 보유한 국민건강보험공단의 빅데이터를 이용하여 수요자 중심의 맞춤형 서비스를 건강정보 전문사이트 건강iN(hi.nhis.or.kr)을 통하여 구축하여 왔다.

본 연구는 건강iN(hi.nhis.or.kr)의 PHR의 기초 연구로 시작되어 『2014년, 2015년, 2016년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』 연구에 이은 후속 4차년도 연구이다. 국민건강보험공단의 빅데이터실에서 보유하고 있는 일반건강검진 DB 및 맞춤형 DB를 통하여 수요자 중심의 맞춤형 서비스를 개발하였다. 질병 예측 프로그램의 추가 업데이트, 건강iN의 마이헬스뱅크에 당뇨 질환 탐재를 위한 알고리즘 개발, 건강노화나이의 등급화 및 국가적정 표준화 분류를 제시하여 현재 이용자의 건강, 질병정보 제공과 함께 질병 발생의 예방과 질병 관리에 도움이 되고자 하였다.

이 프로젝트의 초창기부터 많은 협조를 아끼지 않으신 대한심장학회, 대한당뇨병학회, 대한신경학회 및 국립암센터의 협조에 감사드린다.

본 연구에는 국민건강보험공단 일산병원 가정의학과 조경희 교수, 유보라 전임의, 일산병원 연구소의 김동욱 박사, 한규태 박사, 국립 암센터 조현수 교수, 김열 교수, 박보영 교수, 신미연 연구원, 아주대학교 김대중 교수, 하경화 박사, 명지병원 서용성 교수, 충북대.이재우교수, 메디에이지 연구소 배철영 교수, 강영곤 소장이 참여하였다.

끝으로 본 보고서의 내용은 저자들의 의견이며 보고서 내용상의 하자 역시 저자들의 책임이며 국민건강보험공단 일산병원 연구소의 공식적인 견해는 아님을 밝혀둔다.

2017년 12월

국민건강보험공단 일산병원장 **강 중 구**  
일산병원 연구소 소 장 **장 호 벨**

# 목 차

요 약 .....	1
제1장 서 론 .....	6
제1절 연구 배경 및 필요성 .....	7
제2절 연구의 목적 .....	8
제2장 이론적 고찰 .....	9
제1절 암 질환 예측 .....	10
제2절 당뇨 질환 예측에 따른 처방메세지 개발 .....	18
제3절 심뇌혈관 질환 예측 .....	19
제4절 건강노화나이 .....	21
제3장 연구 내용 및 방법 .....	24
제1절 연구 내용 .....	25
제2절 연구 방법 .....	26
제3절 각 세부 과제별 연구 방법 .....	29
제4장 분석 결과 .....	44
제1절 암 질환 예측 연구 .....	45
제2절 당뇨 질환 예측에 따른 처방메세지 개발 .....	57
제3절 심뇌혈관 질환 예측 .....	67
제4절 건강노화나이 .....	73
제5장 결론 및 제언 .....	82
제1절 결 론 .....	83
제2절 제 언 .....	88
참고문헌 .....	91

## 표목차

<표 2-1> 각 연구에서 사용된 위암 발생 위험 인자 .....	12
<표 2-2> 대장암 예측모형별 위험요인 비교분석 .....	17
<표 2-3> 국내 폐암 예측 모델에서 사용된 risk factor .....	15
<표 2-4> 폐암 예측 모델 변수 목록 .....	17
<표 3-1> 표본코호트DB 세부특성 .....	27
<표 3-2> 건강검진 코호트 DB 보유 정보 주요 내용 .....	28
<표 3-3> 위암 예측 모델 구축시 Outlier 제거 기준 .....	30
<표 3-4> 연도별 문진에 따라 산출 가능한 흡연 변수 항목 .....	34
<표 3-5> 건강인 선정을 위한 생체지표 범위 .....	42
<표 4-1> 위암 예측 모델 분석 대상자의 특성 - 남성 .....	45
<표 4-2> 나이를 보정한 단변량 모델과 다변량 모델 - 남성 위암 모델 .....	47
<표 4-3> 남성, 여성 위암 모델 변수 차이 .....	48
<표 4-4> 나이를 보정한 단변량 모델과 다변량 모델 - 남성 추가 모형 .....	49
<표 4-5> 위암 예측 모델 분석 대상자의 특성 - 여성 .....	50
<표 4-6> 나이를 보정한 단변량 모델과 다변량 모델 - 여성 위암 모델 .....	51
<표 4-7> 남성의 대장암 발생 위험요인별 다변량 cox regression 분석결과 .....	53
<표 4-8> 여성의 대장암 발생 위험요인별 다변량 cox regression 분석결과 .....	54
<표 4-9> 폐암 예측 연구대상자 흡연력 분포 .....	55
<표 4-10> 폐암 예측 흡연 외 고려 변수 .....	56
<표 4-11> 1000,000-5 Person Year 폐암 발생률 .....	56
<표 4-12> Hazard ratio for related factors for incidence diabetes .....	58

<표 4-13> 10년 당뇨 위험 예측도 .....	60
<표 4-14> 당뇨 위험 교정 가능한 요인 분류 설정 .....	63
<표 4-15> 심뇌혈관 질환 발생 여부에 따른 대상군의 특성 .....	67
<표 4-16> 심뇌혈관 질환 발생의 위험 요인 .....	69
<표 4-17> 남녀에서의 새로운 심뇌혈관 통합 예측 모델 .....	71
<표 5-1> 폐암 예측 질병 카테고리 .....	84
<표 5-2> 기존에 발표된 심뇌혈관 예측 모델의 변별력 (C-index) .....	85

## 그림목차

[그림 2-1] 2015년 암종별 사망자수, 통계청 .....	14
[그림 2-2] 2015년 사망 원인 통계, 통계청 .....	19
[그림 2-3] 심뇌혈관 예측모델 위험인자와 예측 결과 .....	21
[그림 3-1] 위암 예측 대상자 선정 순서도 .....	29
[그림 3-2] 연구 DB 구축 및 자료의 보관·사용 관련 체계 .....	32
[그림 3-3] 연구 대상자 선정을 위한 연구흐름도 .....	33
[그림 3-4] 폐암 발생 위험 예측 모델 개발 set 모형도 .....	35
[그림 3-5] 폐암 발생 위험 예측 모델 개발 set 대상자 수 .....	36
[그림 3-6] 당뇨 위험 예측도 구간 설정 .....	38
[그림 3-7] 2016년 건강노화나이 검사항목 .....	41
[그림 3-8] 건강노화나이 자료분석구조 .....	42
[그림 4-1] 남성 위암 모델 calibration 막대 그림 .....	49
[그림 4-2] 여성 위암 모델 calibration 막대 그림 .....	52
[그림 4-3] 당뇨 위험도 개선을 위한 교정 가능한 요인 .....	62
[그림 4-4] 심뇌혈관 통합 예측 모델의 Calibration plot .....	72
[그림 4-5] 심뇌혈관 발생 위험도에 따른 혈중 콜레스테롤 수치의 변화 .....	73
[그림 4-6] 연령에 따른 신장 변화 .....	73
[그림 4-7] 연령에 따른 체중, BMI,허리둘레 변화 .....	74
[그림 4-8] 연령에 따른 혈압 변화 .....	75
[그림 4-9] 연령에 따른 lipid profile 변화 .....	76
[그림 4-10]연령에 따른 간기능 변화 .....	77



[그림 4-11] 연령에 따른 Hgb,,FBS 변화 .....	77
[그림 4-12] 연령에 따른 신기능 변화 .....	78
[그림 4-13] 활동량에 따른 건강노화나이 차이 .....	78
[그림 4-14] 흡연에 따른 건강노화나이 차이 .....	79
[그림 4-15] 음주에 따른 건강노화나이 차이 .....	80
[그림 4-16] 연령대 별 분포: 주민등록나이 vs 건강노화나이 .....	80
[그림 4-17] 건강노화나이에 따른 등급화 예제 .....	81

요약

## 요 약

### 1. 연구 배경 및 목적

본 연구는 『2014년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』 연구에 이은 후속 4차년도 연구로써, 일반건강검진 및 맞춤형 DB를 활용해 질병 예측 프로그램을 개발 및 건강노화나이의 등급화 및 국가적정 표준화 분류하여 현재 이용자의 건강, 질병정보 제공을 통하여 질병 발생의 예방과 질병 관리에 도움이 되고자 한다.

### 2. 연구 내용 및 방법

맞춤형 자료를 사용한 건강검진결과와 산정특례 등록자료를 연계하여 위암예측 프로그램을 신규 개발하였다. 국민건강보험공단 청구자료와 건강검진 자료를 이용한 대장암 발생현황과 위험요인들을 확인하였다.<sup>22</sup> 폐암은 현재 국립암센터에서 진행 중인 폐암 예측 프로그램과 공동협조로 전수조사를 통하여 진행하였다.

2016년 3차년도 연구에서 개발한 당뇨질환 예측 모형에 대한 건강iN의 마이헬스뱅크에 신규 질환 탑재를 위한 알고리즘을 개발하였다. 국민 건강 보험 공단의 건강 검진 자료를 이용하여 비혈액검사적 방법을 통한 심뇌혈관 통합 예측 모델을 개발하였다.

건강노화나이에 사용되는 각 생체지표의 연령별, 성별 국가 현황 및 흡연, 음주, 운동 등의 생활습관과의 연관성 조사하였으며, 기존에 개발한 건강노화나이 모델을 사용한 건강노화나이 등급화를 개발하였다. 이를 사용하여 수검자의 건강노화나이 상태를 제시하여 건강관리를 집중화시킬 수 있는 방안을 제시해 주고자 하였다.

### 3. 연구 결과

2004, 2005년 건강검진 총 수검자를 대상으로 한 위암 예측 모델에 포함된 변수로 남성에서는 (나이-평균나이), (나이-평균나이)<sup>2</sup>, 비만도, 가족력, 식이습관, 음주량(bottle/day), 흡연량(pack/year), 운동 횟수(per week)이 있었으며, 추적기간 10년간 위암 발생을 확인한 결과 남성 5,482,191명중 86,308명(1.57%)에서 위암 발생이 확인되었다. 남성 위암 모델에 대한 C statistics는 모델구축용 데이터에서는 0.868[95% CI: 0.866-0.871], 검증용 데이터에서는 0.863[95% CI: 0.859-0.867]로 나타났다. 이는 개발된 모델의 판별력이 상당히 높음을 보여주지만, calibration 결과는 개발된 모델의 예측값과 실제값에 차이가 있음을 나타내었다. 여성 위암 예측 모델에 포함된 변수로 (나이-평균나이), (나이-평균나이)<sup>2</sup>, 비만도, 가족력, 식이 습관, 음주량(bottle/day), 흡연량(pack/year), 운동 횟수(per week)이며 추적기간 10년간 위암 발생을 확인한 결과 여성 4,443,515명중 35,388명에서 위암 발생이 확인되었다. 여성 모델 평가여성 위암 모델에 대한 C statistics은 모델구축용 데이터에서는 0.803[95% CI:0.797-0.810], 검증용 데이터에서는 0.812[95% CI: 0.803-0.821]로 나타났다. 이는 개발된 모델의 판별력이 상당히 높음을 보여주지만, calibration 결과는 개발된 모델의 예측값과 실제값에 차이가 있음을 나타내었다.

당뇨 위험 예측식을 이용하여, 10년 당뇨 발생 위험 예측도를 산출하였으며, 남자의 50%에 해당하는 당뇨 위험 예측도는 30.9%였으며, 여자는 12.9%였으며, 남자 70% 및 여자 70%에 해당하는 예측도는 각각 47.5%, 21.1%였고, 90%에 해당하는 값은 76.2%, 33.5%였다. 또한 당뇨 위험도 개선을 위한 생활습관 처방 메시지를 개발하였다.

2002, 2003년 건강검진을 받은 건강검진 코호트 대상군에서 총 4,917,150 인-년(person-year)의 추적 기간 동안 남성은 약 7.0%인 17,165명에서 여성은 약 5.9%인 11,238명에서 심뇌혈관질환이 발생하였다. 비혈액검사적 요인을 사용하여 심뇌혈관 통합 예측 모델을 개발하였으며, 남성 모델의 경우 나이와 수축기 혈압, 고혈압 약제 복용 여부, 당뇨, 체질량 지수, 흡연력, 운동습관 및 음주 습관이 최종 모델의 변수에 포함되었으나 여성의 경우에는 음주 습관은 최종 모델에서 배제되었다.

최종 예측 모델의 경우, 남성 심뇌혈관 질환 모델에 대한 C statistics은 모델구축용 데이터에서는 0.760, 검증용 데이터에서는 0.752로 나타났으며, 여성 심뇌혈관 질환 모델에 대한 C statistics은 모델구축용 데이터에서는 0.792, 검증용 데이터에서는 0.788로 나타났다. 남성과 여성 모두 고위험군에서 위험도를 과다 추정하는 경향이 발견되었으나 실제 값의 차이가 크지 않아 용인할 수 있는 수준의 오차를 보인 것으로 판단되었다.

남자에서 연령 증가에 따라 증가하는 생체지표로서는 허리둘레, 수축기혈압, 이완기혈압, 공복혈당, 총콜레스테롤, 저밀도콜레스테롤, 요단백, 지오티가 있었다. 반대로 연령 증가에 따라 감소하는 생체지표로서 신장, 체중, 체질량지수, 혈색소, 중성지방, 고밀도콜레스테롤, 혈중크레아티닌, 사구체여과율, 지피티, 감마지피티가 있었다. 여자에서 연령 증가에 따라 증가하는 생체지표로서는 체질량지수, 허리둘레, 수축기혈압, 이완기혈압, 혈색소, 공복혈당, 총콜레스테롤, 중성지방, 저밀도콜레스테롤, 혈중크레아티닌, 지오티, 지피티, 감마지피티가 있었다. 감소하는 생체지표로서는 신장, 체중, 고밀도콜레스테롤, 요단백, 사구체여과율이 있었다. 남녀 모두 활동량이 많을수록 건강노화나이는 더 좋았으며 건강노화나이의 주민등록 나이와의 차이가 적었다. 남녀 모두 비흡연군에 비해 흡연군에서 건강노화나이가 나쁜 것으로 분석되었다. 남녀 모두 음주군에서 건강노화나이가 나쁜 것으로 분석되었다.

#### 4. 결론 및 제언

본 연구에서는 건강보험공단 건강검진자료를 이용한 인구기반 위암 위험 예측 모델을 개발을 목적으로 과거에 건강검진자료를 기반으로 개발된 위암 위험 모형의 활용성을 점검하였다. 기존 모형은 10년 전 자료를 근거로 생성된 모델이므로 현재의 인구자료에 활용하는데 한계가 있으며, 선행 위암 위험 발생 예측 모델 연구와 동일한 접근 방법으로 최신 자료를 이용하여 10년 위암 발생 예측 모델을 구현 하였다. 향후 건강검진 수검자들의 검진결과를 기반으로 위험인자를 측정하여 위암 위험 발생 예측 모델을 구축하고 이용자로 하여금 건강iN 서비스를 통해 위암 발생 위험요소를 파악하고 건강한 생활습관의 변화를 유도할 수 있도록 하는데 그 의의가 있다고 할 수 있다.

당뇨 위험 예측식을 추정하고 이를 이용하여, 이용자들의 개인의 각 당뇨 위험 요인의 값을 통하여 개인별 10년 당뇨 위험 예측도를 추정하고 제시하였다. 이용자들의 개인의 당뇨 위험도의 교정 가능한 요인을 제시받고, 각 위험 요인별 생활습관 교정 메시지를 제공받을 수 있도록 하였다.

비혈액검사적 방법을 이용한 심뇌혈관 예측 모델은 단순히 비용 절감을 넘어 의료기관에 방문하여 검사를 받거나 건강검진을 받지 않아도 건강인 서비스 이용자들이 손쉽게 본인의 심뇌혈관 발생 위험도를 계산할 수 있도록 하여 서비스 이용의 진입 장벽을 낮추는 효과를 나타낼 것으로 기대한다. 혈액검사 없이 반복적으로 본인의 질병 위험도의 변화를 계산할 수 있게 하여 생활습관 변화에 즉각적인 피드백을 함으로서 보다 적극적인 생활습관 변화를 유도할 수 있는 효과를 나타낼 수 있는 부가적인 효과도 기대한다.

기존에 개발된 건강노화나이 모델을 2014-15년 국민건강보험공단 전수자료에 적용하여, 국가 적정 표준화를 제시하였다. 건강노화나이 등급화를 통해서 이는 추후 "나이"에 따른 상호 소통과 비교 이외에 또 다른 동기부여 수단이나 관리목표로 활용될 수 있을 것이다. 일반 수검자들의 만성질환 예방과 건강 증진에 실제적 도움이 될 수 있는 지표로서, 추후 이는 건강검진 결과를 통해 개인의 건강 상태와 노화 정도를 종합적으로 제시하고 개별 수검자들의 결과를 서로 쉽게 이해하고 소통하며 올바른 생활습관 관리를 위한 구체적 관리목표 제공할 수 있을 것이다.

# 제 1 장

## 서 론

제 1 절 연구의 배경 및 필요성	7
제 2 절 연구의 목적	8

# 제1장 서론

## 제1절 연구 배경 및 필요성

의료 이용자들의 건강에 대한 관심은 점차 증가하고 있으며 자신의 건강과 관련된 요인들을 스스로 결정하고자 하는 욕구가 증가하고 있다. 또한 의료 이용자들의 생활수준이 향상되고 의학기술이 급속하게 발달함에 따라 질병예방중심 의료로의 보건의료 패러다임이 전환되고 있다.

지역단위를 대표하는 만성질환 관리지표에 대한 요구 증가하고, 전 국민 건강검진 관련 정보를 보유한 건강정보 전문사이트 건강iN(hi.nhis.or.kr)의 강점을 살린 수요자 중심의 맞춤형 서비스 제공이 필요하며, 인터넷, TV등 사회에 무분별하게 범람하는 많은 건강정보로부터 의료 이용자를 보호하기 위하여 신뢰할 수 있는 건강·질병정보 제공을 위한 사이트를 구축해야할 필요성이 제기되고 있다. 현재 국민건강보험공단에서 운영하고 있는 건강iN에서는 건강정보 전문사이트로서 공단이 보유하고 있는 건강자료를 이용한 개인별 맞춤형 건강정보 제공을 하고 있으며 개인의 건강위험도 평가 후 건강개선자료를 제공하는 등 맞춤형 정보를 제공하고 있다.

『2014년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 과거 건강iN 서비스의 여러 문제점들을 제시하였고, 이를 바탕으로 『2015년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 의료 이용자들에게 개인의 건강검진 정보를 통한 통합 처방 처방메세지를 제공함으로써 통합된 개인 건강기록서비스(PHR, personal health record)를 구축하였고 또한 새롭게 뇌졸중 위험 예측 모델 과 심장질환 위험 예측 모델을 개발하여 건강iN 서비스의 ‘My Health Bank’의 이름으로 탑재하여, 의료 이용자들의 건강위험요소를 파악하여 생활습관을 교정할 수 있도록 하였다. 『2016년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 이전



개발한 당뇨 위험 예측 모델 개발 및 타당도 확인, 뇌졸중 및 심장질환 위험 예측 프로그램 타당도 및 업데이트, 건강나이 모델 개발 및 검증한 바 있다.

개인건강기록서비스(PHR)을 통해 능동적으로 개인의 건강정보를 보고 행동변화를 유도할 수 있을 것이다. 국민 개개인의 행동변화를 유도함으로써 예방중심의 의학 패러다임으로 유도를 하며, 궁극적으로는 전체 의료비 절감 효과를 볼 수 있을 것이다.

## 제2절 연구의 목적

본 연구는 『2014년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』 연구에 이은 후속 4차년도 연구로써, 일반건강검진 DB 및 맞춤형 DB를 통하여 질병 예측 프로그램을 추가 업데이트 및 건강노화나이의 등급화 및 국가적정 표준화 분류하여 현재 이용자의 건강, 질병정보 제공을 통하여 질병 발생의 예방과 질병 관리에 도움이 되고자 한다.

- 1) 암질환 예측 모델 개발
- 2) 당뇨 질환 위험 예측 모델 고도화 및 건강인 웹페이지 탑재 알고리즘 개발
- 3) 심뇌혈관 질환 예측 모델 고도화
- 4) 각 생체지표의 연령별, 성별 국가 현황과 생활습관 항목과의 연관성 조사 및 건강노화나이 고도화 개발

# 제2장

## 이론적 고찰

제 1 절 암 질환 예측	10
제 2 절 당뇨 질환 예측에 따른 처방메세지 개발	10
제 3 절 심뇌혈관 질환 예측	19
제 4 절 건강노화나이	21

---

## 제2장 이론적 고찰

### 제1절 암 질환 예측

#### 1. 위암 예측 모형 문헌고찰

Eom et al은 건강보험공단에서 연구목적으로 공개한 국민건강보험공단 1차 검진, 문진 자료를 기반으로 1996~1997년 동안 건강검진을 받은 1,391,889명을 대상으로 나이, BMI, 가족력, 식사 횟수, 염분 섭취량, 식이 습관, 육류 섭취량, 음주량(g/day), 흡연량(pack/day), 육체활동을 예측인자로 고려하여 위암 예측모형을 구축하여 2015년에 발표하였다. 해당 연구에서 남성, 여성의 ROC AUC값이 각각 0.764(95% CI, 0.760-0.768), 0.706(95% CI, 0.689-0.715)으로 좋은 설명력은 보여 주었다. 하지만 이 모형은 10여년 전 자료를 사용한 연구이므로 현재 적용하기에는 한계가 있었다.<sup>1</sup>

Park et al은 2008년 1월~2013년12월 동안 강남 세브란스 병원 건강 증진 센터에서 위 내시경 검사를 받은 사람 60,261명을 대상으로 cross-sectional 조사를 진행하였다. 연구 참여자의 임상 데이터는 전자 의료 기록을 검토하여 나이, 성별, BMI, 음주량, 식습관, 당뇨유무, 흡연, Helicobacter pylori, 위축성 위염, 장상피화생을 위험 인자로 고려하여 위암과의 연관성을 확인하였다. 내시경 biopsy를 통해 확인 된 위암은 WHO 분류에 따라 보고되었고, 조직학 특성은 로렌 분류에 따라 평가되었다. 추가적으로 위암으로 진행되는 전암 상태에서의 위험 인자도 고려하였다. 하지만 설문지에는 섭취 한 음식에 함유된 나트륨의 양이나 농도는 포함되어 있지 않는 한계가 있었다.<sup>2</sup>

Lee et al은 대만의 국민 건강 보험 연구 데이터베이스(National Health Insurance Research Database)를 기반으로 1997년 1월 1일~2004년 12월 31일 사이에 278,898명의 소화성 궤양 진단 초진 환자를 대상으로 연구를 진행하였다. 나이,

성별, 위궤양, 소화성 궤양 합병증, Helicobacter pylori 제균, 비스테로이드성 항염증제 사용, 감시 내시경 검사 및 동반질병과 같이 위암 위험과 관련이 있는 인자들을 사용하여 소화성 궤양 환자를 위한 맞춤형 위험 예측 모델 개발하였으며, 개발된 노모그램의 AUC값은 0.78이었다.<sup>3</sup>

Wang et al은 1993-1998년 동안 등록된 45-74세의 63,257명 남녀의 전향적 코호트인 싱가포르 중국 건강 연구를 사용하여 여러 가지 생활 습관 요소와 위선암 발생 위험 사이의 관련성을 조사하기 위한 연구를 수행하였다. 위암의 6가지 개선 가능한 위험 인자인 흡연, 음주, 비만, 신체 활동, 염분 섭취, 식이 습관을 사용하여 종합 보호 생활 습관 점수를 계산하였다. 추가적으로 위선암 위험과 관련하여 종합 보호 생활 습관 점수와 H. pylori 감염 상태의 연관성을 조사하기 위해, 환자-대조군 연구가 522명(위암 133 예와 개별 대조 대조군 389)을 대상으로 실시되었다. 결과적으로 건강한 종합 생활 습관 점수는 위선암의 위험 감소와 유의한 관련이 있었다.<sup>4</sup>

Buckland et al은 유럽 10개국의 23개 센터에서 진행 중인 코호트 연구에서 1992-2000년 동안 등록된 461,550명을 대상으로 흡연, 음주, 식이습관(지중해식)을 사용한 건강한 생활 습관 지수를 계산해 중앙 부위 및 조직학적 유형에 따른 위암 발생 위험 확인하는 연구를 수행하였다. 인구 집단의 위험도 계산 결과, 금연, 절주, 건강 식단 섭취 및 정상 체중을 유지하는 등 건강한 생활 습관을 수행 할 경우 모든 위암의 18.8%와 위분문부(Cardia)경우의 62.4%가 예방 될 수 있다는 결과를 보여주었다.<sup>5</sup>

<표 2-1> 각 연구에서 사용된 위암 발생 위험 인자

	Forman et al,2006	Tran (코호트) et al, 2005 중국	Charvat (코호트)et al, 2015 일본	Lee (대만코호트) et al, 2015	Buckland (코호트) et al, 2015	Wang Et al,2017	Park Et al ,2016	Eom, (코호트) Et al 2016
Age		Yellow	Teal	Blue	Cyan		Red	Green
Gender		Yellow	Teal	Blue	Cyan		Red	
Height		Yellow						
DM		Yellow					Red	
BMI	Dark Blue	Yellow			Cyan	Orange	Red	Green
Smoking			Teal		Cyan	Orange	Red	Green
Alcohol	Dark Blue		Teal		Cyan	Orange	Red	Green
physical activity					Cyan			Green
Education					Cyan			
가족력	Family Hx of cancer	Yellow						Green
식이	Meat consumption	Yellow						Green
	Meal regularity							Green
	Meal preference	Fruits,vegetary	Meat, eggs, fresh fruits	Teal		Energy intake	Orange	Green
	Salt preference			Teal			Orange	Red
Occupation	Dark Blue							
Pernicious anemia	Dark Blue							
Peptic ulcer disease	Dark Blue			Blue				
Prior Hx of gastric surgery	Dark Blue							
NSAID Use	Dark Blue			Blue				
H.Pylori infection/ eradication				Blue			Red	
Endoscopy Surveillance				Blue				
Atrophic gastritis			Teal				AG,IM	

## 2. 대장암 발생 위험 예측

대장암 발생 예측에 대한 <표 2-2>와 같이 여러 예측 모형들이 존재하며 모형 적합도가 70% 내외로 나타나고 있다. 대장암 발생의 위험 요소로는 연령, 성별, 흡연, 가족력, 음주, 비만 등의 요소가 공통적으로 지적되고 있으며 자료가 있는 경우에는 이전의 대장내시경 시행 결과의 폴립유무 등이 포함되어 있음. 이 외에도 운동량, 식습관, 당뇨 등의 요소가 포함된 모델들이 있었다.

국내에서는 Shin et al 이 2014년에 5년간 암이 발생할 위험을 예측하는 모형을 남녀에서 개별적으로 개발하여 연구하였다.<sup>6</sup> 30세에서 80세 사이의 인구를 대상으로

1996년에서 1997년에 건강검진을 받은 846,559명의 남성 및 479,449명의 여성에서 향후 colorectal cancer의 발병 여부를 Korean Central Cancer Registry(KCCR)database 등록자료로 확인하였다. 이후, 모형을 개발하였으며 1998년과 1999년 사이에 건강검진을 받은 547,874명의 남성과 415,875명의 여성에서 적합도를 평가하였다. 해당 모형은 proximal/distal colon cancer 및 rectal cancer에 대하여 개별적 모형으로 개발되어 있다. 또한 남성과 여성에서 colorectal ca, Lt/Rt colon ca, colon ca, rectal ca등의 5가지 모형을 각각 개발하였으며 C-statistic으로 평가한 모형의 예측력은 0.69에서 0.78사이의 값으로 확인되었다. 남성에서는 모형에 의해 예측된 발병률과 실제 결과가 매우 비슷하게 나타났으나 여성에서는 유의한 예측력을 보이지 않았다.

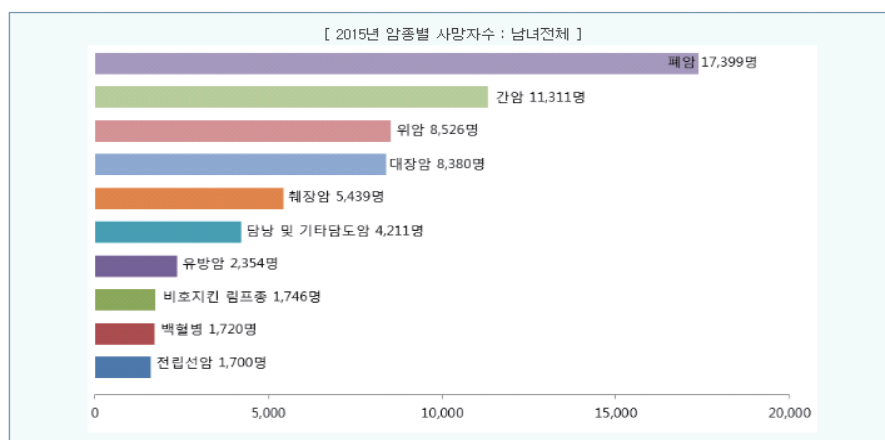
<표 2-2> 대장암 예측모형별 위험요인 비교분석

	Cai et al,2012	Yeoh et al,2011	Ma et al,2010	Wei et al,2009	Freedman and Park et al, 2009	Kastrinos et al, 2009	Driver et al,2007	Lin et al, 2006	Imperial e et al, 2003	Levitzy et al,2011	Korean
Age											남여
Gender											
Smoking											
DM											
Green vegetables											
Picked food											
Fried food											
FDR with CRC											
SDR with CRC											
BMI											남
Alcohol											남frequency & amount
Physical activity											
Meat consumption	White meat			Red or processed meat							여
Folate											
Sigmoidscopy results											
Colonoscopy results											
Hx of polyps											

Relative with CRC						more than 3 relatives						
Aspirin/NSAID Use												
Estrogen status(women only)												
CRC or polyp diagnosed before age 50												
Height												여
Fasting glucose												
Total cholesterol												남
Family Hx of cancer												남

### 3. 폐암 발생 위험 예측

2017년도 국가암정보센터에 의하면 연간 24,027명의 폐암환자가 발생하며 전체 암 사망의 22.6%를 차지한다고 발표하였다.



[그림 2-1] 2015년 암종별 사망자수

2009-2013년 폐암의 5대 상대 생존율은 남성에서 20.5%, 여자 30.6%를 보였다. 병기별 5년 상대 생존율에서는 국한병기 53.9%, 국소진행 26.7%, 원격전이 5.1%를 보이며 국한된 병소 단계에서 조기발견이 중요함을 알 수 있다.

Park et al 연구에 따르면 국내에서 처음으로 시행된 폐암 위험 예측 모델 연구는 아시아인 대상으로 한 첫 폐암발생 예측 모형으로 130만명 한국남성을 대상으로 8년간 추적관찰을 하였다.<sup>7</sup> 모형에서 <표 2-3>와 같이 연령, 체질량지수, 흡연여부,

흡연량, 흡연 시작나이, 혈당, 신체활동 등의 폐암발생의 기여도를 분석한 것이며, 예측력은 0.87로 높은 값을 나타냈으나, 남성에서만 연구되어 향후 여성 대상 모형 필요하며 폐암에 있어 유의한 관계가 있는 가족력, 간접흡연, 환경오염, 직업의 특이성 등이 향후 추가로 고려될 필요가 있다.

<표 2-3> 국내 폐암 예측 모델에서 사용된 risk factor <sup>7</sup>

Risk factor	$\beta$	HR* (95% CI)	p-value	p for trend
Age-Mean <sub>age</sub> , years	0.1668	1.18 (1.18-1.19)	<0.0001	
(Age-Mean <sub>age</sub> ) <sup>2</sup> , years <sup>2</sup>	-0.0020	1.00 (1.00-1.00)	<0.0001	
<b>Smoke</b>				
Never		1.00 (reference)		<0.0001
Past	0.4180	1.52 (1.41-1.64)	<0.0001	
Current, <0.5 pack/day	0.4444	1.56 (1.42-1.71)	<0.0001	
Current, 0.5-0.99 pack/day	0.9414	2.56 (2.37-2.78)	<0.0001	
Current, $\geq 1$ pack/day	1.3889	4.01 (3.68-4.37)	<0.0001	
<b>Age at smoking initiation</b>				
Age $\geq 40$		1.000 (reference)		<0.0001
30 $\leq$ Age <40	0.2194	1.25 (1.16-1.34)	<0.0001	
19 $\leq$ Age <30	0.2809	1.32 (1.23-1.43)	<0.0001	
16 $\leq$ Age <19	0.5249	1.69 (1.4-2.04)	<0.0001	
Age <16	0.7120	2.04 (1.46-2.84)	<0.0001	
<b>BMI, kg/m<sup>2</sup></b>				
<18.5	0.3306	1.39 (1.28-1.51)	<0.0001	<0.0001
18.5-22.9		1.00 (reference)		
23.0-24.9	-0.2468	0.78 (0.74-0.82)	<0.0001	
$\geq 25.0$	-0.3386	0.71 (0.68-0.75)	<0.0001	
<b>Physical activity</b>				
No		1.000 (reference)		<0.0001
Light	-0.0909	0.91 (0.86-0.97)	0.0029	
Moderate	-0.1412	0.87 (0.83-0.91)	<0.0001	
Heavy	-0.0521	0.95 (0.89-1.02)	0.1431	
<b>Fasting glucose level, mg/dL</b>				
<126		1.000 (reference)		
$\geq 126$	0.0792	1.08 (1.01-1.16)	0.0201	0.0201

지난 『2015년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서 폐암 예측 프로그램을 개발한 바 있다. 2002-2003년도에 검진 및 문진 내역이 있는 20세 이상의 검진 대상자로 폐암 내역을 10년간 추적 관찰하였다. 연구 대상자 총 3,806,806명(남자 2,376,129명, 여자 1,430,677명) 중 폐암 발생은 남자는 1.72%인



40,917명, 여자는 1.10%인 15,764명이 발생하였다. 남자인 경우 위험인자는 연령, 연령 제곱, 고혈압 과거력, 뇌졸중 가족력, 운동, 음주, 총콜레스테롤, 흡연, 당뇨 및 체질량지수로 구성되었으며 여자인 경우 위험인자는 연령, 연령 제곱, 과거력 심장병, 음주, 흡연 및 체질량지수로 구성되었다. 예측력은 남성 AUC 0.80(0.77,0.83), 여성 0.76(0.72,0.79)로 분석되었다.

폐암 예측 모델 개발 관련하여 해외 연구 동향을 살펴보면, 폐암 예측 모델 개발에 있어서 흡연력 외에 다른 위험요소들을 포함하고 있는 모델들이 개발되어 있다<표 2-4>. 2016년 3월 발표된 Risk Prediction Models for Lung Cancer에 대한 systematic review에 따르면<sup>8</sup> 31개의 문헌에서 25개의 risk model을 비교하였는데, 11개의 문헌은 역학적 요인(epidemiological factors)에 초점을 맞추어 보았으며 다른 11개의 문헌은 혈액검사나 scan같은 임상적인 평가에 초점을 두었고, 8개의 문헌은 2 단계의 클론 확장(clonal expansion)에 초점을 두었다. 논문에서는 각 risk model들이 각기 다른 인구집단을 대상으로 타당도 검사를 다양한 방식으로 했기 때문에 이를 비교하기 위해선 같은 dataset으로 각 모델을 비교할 필요가 있음을 이야기하였으나 그럼에도 불구하고 PLCOM2012와 Hoggart model이 전반적으로 훌륭한 성과를 보였다.

<표 2-4> 폐암 예측 모델 변수 목록<sup>8</sup>

Risk Factors	Bach <sup>13</sup>	Liverpool Lung Project <sup>14</sup>	Spitz <sup>15</sup>	African-American <sup>16</sup>	PLCO <sup>17</sup>		PLCO <sub>M2012</sub> <sup>18</sup>	Hoggart <sup>19</sup>
<b>Personal Factors</b>								
Age	Y	Y	Y	Y	Y	Y	Y	Y
Sex	Y	Y	Y	Y	N	N	N	N
Ethnicity	N	N	N	N	N	N	Y	N
Body mass index	N	N	N	N	Y	Y	Y	N
Subject to x-rays	N	N	N	N	Y	Y	N	N
Education (levels)	N	N	N	N	Y	Y	Y	N
Previous malignant tumor	N	Y	N	N	N	N	Y	N
<b>Smoking History</b>								
Smoking status	N	N	Y	Y	Y	Y	Y	Y
Start age	N	N	N	N	N	N	N	Y
Cessation age	N	N	Y	Y	N	N	N	N
Smoking duration	Y	Y	N	N	Y	Y	Y	Y
Cigarettes per day	Y	N	N	N	N	N	Y	Y
Pack-years	N	N	Y	Y	Y	Y	N	N
Quit duration	Y	N	N	Y	N	Y	Y	N
Environmental smoke	N	N	Y	N	N	N	N	N
<b>Family History of Cancer</b>								
Cases of any cancer	N	N	N	N	N	N	N	N
Age of onset for any cancer	N	N	N	N	N	N	N	N
Cases of smoking-related cancer	N	N	Y	N	N	N	N	N
Age of onset for smoking cancer	N	N	N	N	N	N	N	N
Cases of lung cancer	N	Y	Y	N	Y	Y	Y	N
Age of onset of lung cancer	N	Y	N	N	N	N	N	N
<b>Exposures and Lung Conditions</b>								
Asbestos exposure	Y	Y	Y	N	N	N	N	N
Dust	N	N	Y	Y	N	N	N	N
Hay fever	N	N	Y	Y	N	N	N	N
Asthma	N	N	N	N	N	N	N	N
Emphysema	N	N	Y	N	N	N	N	N
COPD	N	N	N	Y	Y	Y	Y	N
Pneumonia	N	Y	N	Y	N	N	N	N
<b>Applying the Model Information</b>								
Applicable to never smokers	N	Y	Y	Y	Y	N	N	N
Applicable to former smokers	Y	Y	Y	Y	Y	Y	Y	Y
Applicable to current smokers	Y	Y	Y	Y	Y	Y	Y	Y
Model age restrictions	50-75	40-80	≥20	≥20	N	N	N	≥35
Model smoking restrictions	≥30 PY	N	N	N	N	N	N	N
Model predicts risk of incidence	Y	Y	N	N	Y	Y	Y	Y
Model predicts survival	Y	N	Y	Y	N	N	N	Y
Length risk length, years	1-10	5	≥1	5	9	9	6	≥1
Model formula printed	Y	Y	Y	Y	Y	Y	Y	Y

## 제2절 당뇨 질환 예측에 따른 처방메세지 개발

### 1. 당뇨병 유병 증가

세계보건기구의 2016년 보고서에 따르면, 2015년에 150만명이 당뇨 관련 질환으로 사망하였으며, 1980년에 비하여 2014년에는 3배 이상 유병이 증가해 44,200만명을 넘어, 전세계 인구 11명중에 1명은 당뇨라고 보고하였다(Global report on diabetes WHO, 2016).<sup>9</sup> 국제당뇨연합(IDF)에서도 2015년 41,500만명에서 2040년 64,200만명으로 증가할 것으로 예측하고 있다(IDF, IDF diabetes atlas 2015).<sup>10</sup> 한국당뇨학회(KDA)의 Fact sheet에서 30세 이상 우리나라 당뇨병 유병률 또한 2012년 10.1%에서 2016년 13.7%로 증가하였다고 보고되었고, 다른 연구에서도 2001년 8.6%에서 2013년 11.0%로 증가하였다고 보고되었다(KH Ha & DJ Kim, Endocrinol Metab 2015).<sup>11,12</sup> 또한 당뇨병 약제 청구 인원 규모는 2015년 약 30만명에서 2020년 약 36만명으로 증가로 예상 하였다(SO Song 등, Endocrinol Metab 2016).<sup>13</sup>

당뇨병의 치료와 합병증의 예방 및 치료에 드는 비용의 증가와 합병증 등으로 인한 경제적 손실로 세계보건기구(WHO)가 중심이 되어 선진국을 비롯한 세계 여러 나라에서 당뇨병 예방에 노력을 기울이고 있으며, 특히 당뇨병 환자의 조기 선별을 통한 일차 예방의 중요성이 높아지고 있다.<sup>14</sup>

### 2. 선행 당뇨 질환 예측 모형

기존 2016년 연구에서 2002년 및 2003년 건강검진 수검자를 대상으로 10년 동안 추적 관찰하였으며, 기존 위험요인과 당뇨병 발생과의 관련성을 확인하기 위해 Cox 회귀분석을 활용하여 당뇨 질환 예측 모형을 개발한 바 있다.

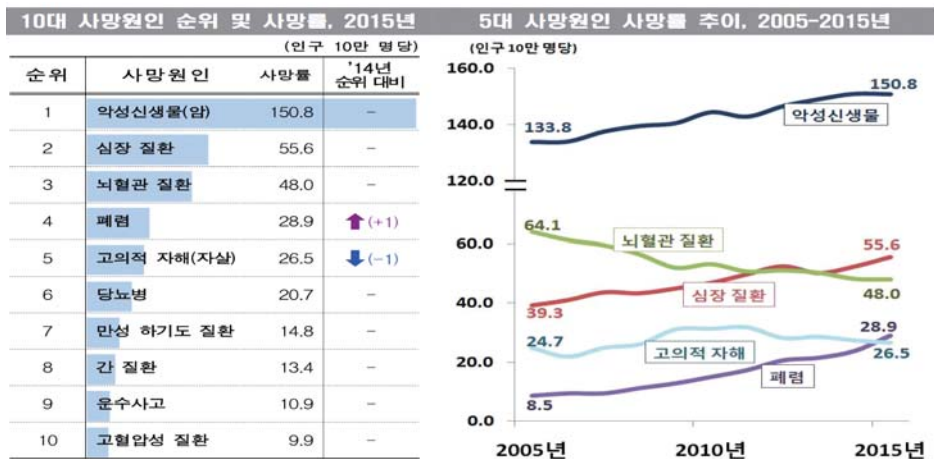
### 3. 공단자료를 이용한 당뇨 예측 모형 및 알고리즘 개발

기존 선행 당뇨 위험 예측 연구 모델은 단면연구를 활용하여 개발하였거나 일개 및 한정된 지역의 연구집단을 대상으로 개발된 연구 결과를 우리나라 전체 국민으로 일반화하여 적용 및 활용하기에는 제한점이 있다. 대국민 건강 정보 서비스를 제공하기 위하여, 국민을 대표할 수 있는 대규모 인구 코호트 자료인 국민건강보험 건강검진

코호트 DB를 통해 당뇨 위험 예측 모형을 개발할 필요성이 제기되었다. 국민건강보험공단의 건강iN 웹사이트에서 뇌졸중 위험예측모델과 심장질환 위험예측모델을 개발하여 알고리즘을 개발하여 탑재했으며, 이용자들에게 맞춤형으로 건강위험 메시지 및 건강정보를 제공하고 있다. 당뇨 질환 예측 연구도 이와 발맞추어, 당뇨 질환 위험 예측 프로그램도 건강 iN 웹페이지에 구현 가능하도록 하는 알고리즘 개발이 필요하다.

### 제3절 심뇌혈관 질환 예측

심뇌혈관 질환(cardiovascular disease)은 우리나라를 포함한 전세계에서 공중보건학적 중요도가 가장 높은 질환 중의 하나이다. 우리나라에서는 통계청의 2015년 사망 통계에 따르면 심장질환과 뇌혈관질환은 전체 사망원인 2위와 3위를 차지하고 있으며, 질병관리 본부에서 발간한 만성 질환 FACTBOOK에 따르면 2014년 기준 순환기계 질환 진료비는 7조 1천억원으로 암 진료비(4조 5천억원)를 상회하는 것으로 알려져 있다.



[그림 2-2] 2015년 사망 원인 통계, 통계청

이러한 심뇌혈관 질환의 발생은 개인의 보건학적 특성에 따른 위험 인자의 보유 여부에 따라 발생률의 차이가 있는데, 심뇌혈관 질환 발생의 위험인자로서는 고혈압, 당뇨병, 고지혈증, 흡연, 비만 및 낮은 신체적 활동 등이 있으며 이는 생활 습관의

교정에 의하여 일정 부분 예방이 가능한 것으로 알려져 있다. 따라서 심뇌혈관 질환은 적절한 치료와 더불어 생활 습관 교정 및 위험 요인의 치료에 의한 예방의 중요성이 매우 강조되는 질환이다.

심뇌혈관 질환의 예측 모델은 이미 전 세계적으로 여러 가지 모델이 개발되어 사용되고 있다. 특히 최근에 있어서는 단순히 질병 발생의 위험도에 대한 정보를 제공하는 것을 넘어 예측된 위험도에 따라 치료 방침을 결정하거나 생활습관 교정의 동기 부여 및 목표 제시를 위한 도구로 사용하는 등 임상적인 활용도가 점점 높아지고 있다.<sup>15,16</sup> 대표적인 모델로 미국 내 전향적 코호트 연구에 기반을 두고 개발된 통합 코호트 모델(Pooled cohort equation)을 들 수 있는데, 현재 미국 심장 학회의 고지혈증 치료 지침은 예측된 심뇌혈관 질환 발생 위험도에 따라 항고지혈증 치료제의 사용 여부를 결정하도록 하고 있다.<sup>17</sup> 또한 기존 연구에 따르면 심혈관 질환 예측 모델을 사용하여 질병발생 위험도를 환자에게 제공하고 이를 기반으로 하여 생활습관 지도를 했을 경우 더욱 효과적으로 심혈관 질환 위험요인을 통제할 수 있는 것으로 알려져 있어 위험요인을 가진 모든 환자에게 정기적으로 질병 예측 모델을 이용한 위험도를 측정하여 정보를 제공하도록 권고하고 있다.<sup>18</sup> 따라서 정확한 심뇌혈관 질환의 위험도를 개개인에게 제공하고 이를 바탕으로 생활 습관 개선 및 적절한 의료 서비스 이용을 유도하는 것은 심뇌혈관질환 예방을 통한 공중보건향상과 비용절감에 효과적일 수 있다.

본 연구자들은 이러한 국제적인 추세에 맞추어 지난 건강인 연구에서 국민건강 보험공단의 건강검진 수검자의 데이터를 바탕으로 뇌혈관 질환 및 심혈관 질환을 예측할 수 있는 질병 위험도 예측 모델을 각각 개발하였으며 이를 바탕으로 생활 습관 교정을 위한 처방 메시지 등을 개발하여 건강인 홈페이지를 통하여 제공하고 있다. 하지만, 기존 연구에서의 몇 가지 제한점이 있어 이를 개선할 필요성이 있다. 첫째로 현재의 모델은 뇌혈관 질환과 심혈관질환을 따로 분류하여 각각의 예측 모델을 제공하고 있는데 이것은 중복 예측으로 개인에게 두 가지 위험 예측값을 제공하게 되고 이로 인하여 불필요한 정보 제공의 혼선을 가져올 수 있다. 심뇌혈관질환은 동맥경화증이라는 공통의 발생 기전을 공유하는 질환으로 위험요인이 유사하며 현재 널리 사용되는 다수의 위험 예측 모델들은 두 질환을 하나로 묶어서 예측하고 있다.

이전에 발표된 심뇌혈관 예측 모델들의 위험인자와 예측 결과의 요약본을 [그림 2-3]에 도시하였다.<sup>18</sup>

따라서 본 연구에서도 국제적인 경향에 맞추어 질환 예측 모델을 개발할 필요가 있으며 이러한 방식은 향후 본 연구의 모델을 다른 예측 모델과 비교하여 평가하기에 유리할 수 있다. 둘째로 기존 연구에 사용되는 연구 대상의 자료에 고밀도 지방산 콜레스테롤을 포함한 혈액 검사에서 획득할 수 있는 주요 인자가 빠져 있는 점이다. 이는 기존 모델의 예측 정확도를 낮추는 한 요인이 되었다. 또한 기존 예측 모델은 혈액검사를 하지 않고서는 질병위험도를 계산할 수 없어 생활 습관의 교정에 따른 혈압이나 체중의 변화와 같은 비혈액검사적 위험요인의 변화에 즉각적으로 대응하여 질병 위험도를 계산할 수 없다. 따라서 혈액 검사 결과의 예측 모델 반영은 현재로서는 예측력의 유의미한 증가없이 모델이 단기간의 생활 습관 교정에 대한 피드백을 할 수 없게 하는 요인이 되었다. 기존 생활 습관교정에 대한 임상지침에서는 단기적이고 가시적인 목표를 환자에게 제공하여 적극적으로 환자의 동기 부여를 유도하도록 권고하고 있다. 따라서 금번 연구에서는 이러한 제한점을 보완하여 심뇌혈관 질환의 통합 예측 모델을 개발하였으며 건강인 서비스 이용자의 서비스 이용의 진입 장벽을 낮추고 반복적인 위험도 측정을 통하여 보다 적극적으로 동기 부여를 할 수 있도록 혈액검사 결과를 배제한 모델을 개발하였다.

## 제4절 건강노화나이

### 1. 생체지표 biomarker

#### 1) 연구배경

같은 연령대라 하더라도 늙어 보이는 사람이 있는가 하면 나이보다 어려 보이는 사람이 있다. 외관상으로 뿐 아니라 기능적인 측면에서도 어떤 사람은 신체적, 정신적 기능을 잘 유지하고 건강하며 사회적으로 왕성한 활동을 하는 사람이 있는 반면 그렇지 못한 사람도 있다. 전자의 사람들을 "생물학적으로 젊다" 라고 하고 후자의 사람들을 "생물학적으로 늙었다"고 할 수 있다. 이처럼 생물학적으로 "젊다" 혹은 "늙었다" 를 객관적으로 측정하기 위해 건강과 노화와 관련된 다양한 생체지표를 종합하여 제시하는 건강노화나이(혹은 생체나이)에 대한 많은 연구들이 지금까지 진행되어 왔다. 또한

Risk Score		Risk Factors/Covariates Included															CVD Events								
Study Group	Study and Region	Data Source	Publication Year	Age	Sex	Total Chol	LDL Chol	HDL Chol	Systolic BP	BP Rx	Diabetes	HbA1c*	Smoking	Family Hx CVD†	Body Mass Index	Social Region	Coronary Revasc	Angina Pectoris	Unstable Angina	Myocardial Infarct	CHD Death	Stroke	Stroke Death	Cardiac Failure	TIA
Framingham CHD <sup>27</sup>	Framingham MA, USA	EAF, EAM	1998	x	x	x	x	x	x	x	x		x					x	x	x	x				
ATP III <sup>28</sup>	Framingham MA, USA	EAF, EAM	2001	x	x	x	x	x	x	x			x							x	x				
Framingham Global <sup>29</sup>	Framingham MA, USA	EAF, EAM	2008	x	x	x	x	x	x	x			x							x	x	x	x	x	x
PRO-CAM <sup>30</sup>	Muenster, Germany	EM	2002	x		x	x	x	x	x			x	x						x	x				
QRISK <sup>31</sup>	QRESEARCH, United Kingdom	EF, EM	2007	x	x	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x
Reynolds Men <sup>32</sup>	Phys Health Study USA	EAF	2008	x		x	x	x	x				x	x			x			x	x	x	x		
Reynolds Women <sup>33</sup>	Women's Health Study USA	EAM	2007	x		x	x	x	x				x	x	x		x			x	x	x	x		
EURO-SCORE <sup>34</sup>	12 cohorts Europe	EF, EM	2003	x	x	x	x	x	x				x				x				x			x	
Pooled Cohort (current)	CARDIA, Framingham, ARIC, CHS, USA	EAF, EAM		x	x	x	x	x	x	x			x							x	x	x	x		

[그림 2-3] 심뇌혈관 예측모델 위험인자와 예측 결과<sup>18</sup>

새로운 4차 산업혁명시대와 함께 평균수명 증가에 따른 인구 고령화, 잘못된 생활습관에 따른 생활습관질환의 증가, 모바일 시대에 따른 원활한 소통의 필요성 등이 강조되며 의료 환경이 급격하게 변화하는 시점에서 적극적으로 변화하고 적응하려는 노력이 필요하다.

## 2) 개념과 의미

건강노화나이(혹은 생체나이)를 구성하는 생체지표(biomarker)는 1980년대 이후 많은 연구들이 진행되었다. 2000년 이전까지 생체지표는 노화에 따른 변화를 잘 반영할 수 있는 지표로서 노화생체지표(Biomarker of aging)를 주로 의미하였다. 2001년 이후 미국노화연구소(NIA)에서 생체지표 개념을 보다 넓게 정의하였다. 즉 생체지표는 정상노화뿐만 아니라, 질환에 따른 변화 혹은 치료에 따른 약물학적 변화를 객관적으로 측정하고 평가할 수 있는 지표로 보다 포괄적 정의를 제시하였다. 2007년에는 미국 심장, 폐, 혈액연구소(NHLBI)에서 생체지표에 유전적 개념을 포함하여 제시하였다<sup>19</sup>. 따라서 현재에는 생체지표가 매우 폭 넓게 정의되고 활용되고 있으며 일부 연구자들은 좁은 의미의 노화생체지표(Biomarker of aging) 이외에

다른 생체지표들을 생체측정지표(biomeasure marker)라고 구별하여 사용하기도 한다.

일반적으로 노화생체지표(Biomarker of aging)는 건강노화나이(혹은 생체나이)를 구성하는 기초 지표로 검사 시점에서 개인의 환경적 영향과 유전적 영향을 모두 반영한다고 알려져 있다. 따라서 건강 상태와 노화 정도를 정확히 반영하는 노화생체지표(Biomarker of aging)를 바탕으로 개발된 건강노화나이는 사망위험과 잔여수명을 평가할 수 있고 노화와 연관된 질환이나 기능저하를 평가할 수 있는 종합지표로 알려져 있다.<sup>20</sup>

### 3) 선정기준

기존 연구에서 생체지표는 노화에 따른 생리적 변화와 노화 관련 질환을 잘 반영하며 검사방법이 쉽고 절차가 용이해야 한다고 한다. 또한 신체 위해가 없고 최소한의 침습적 방법으로 시행할 수 있어야 하며 짧은 기간 내의 미세한 변화를 잘 반영할 수 있는 등의 여러 선정기준들이 제시되었다.<sup>21</sup> 그 외의 선정기준으로는 반복 측정 시 재현성이 높고 건강 상태 혹은 질병 예방에 중요한 지표로서 장수와 연관성 등이 알려져 있다. 그러나 이 같은 모든 기준들에 부합되며 실제 활용할 수 있는 범용적 생체지표는 아직까지 없는 것으로 알려져 있다. 따라서 개별 연구목적과 활용 목적에 적합한 생체지표를 선정하고 이를 기초로 건강노화나이(혹은 생체나이)를 개발하는 것이 필요하다.

## 2. 건강노화나이와 생체지표

본 연구에서 적용된 건강노화나이는 국민건강보험공단 건강검진 신체계측 항목과 혈액검사 항목을 기초로 개발 되었다. 건강노화나이는 동일한 연령대의 사람들과 비교하여 전반적인 건강 상태와 노화 정도를 이해하기 쉬운 “나이” 로 제시한다. 기존 2016년 연구에서는 건강노화나이에 따른 사망률 및 주요 질환 유병률과 발생률에 대한 분석을 시행하였다. 2017년 연구에서는 건강노화나이를 구성하는 생체지표들의 노화에 따른 변화와 생활습관에 따른 차이 분석을 통해 국가 적정 표준화를 진행하였다. 또한 수검자들이 보다 쉽게 개인의 건강과 노화 상태를 인식하고 소통할 수 있도록 건강노화나이 등급화를 제시하였다.



# 제3장

## 연구 내용 및 방법

제 1 절 연구 내용	25
제 2 절 연구 방법	26
제 3 절 각 세부 과제별 연구 방법	29

# 제3장

## 연구 내용 및 방법

### 제1절 연구 내용

#### 1. 암 질환 예측 프로그램 개발(위암, 폐암, 대장암)

올해 새롭게 진행하는 연구로서 맞춤형 자료를 사용한 건강검진결과와 산정특례 등록자료와 연계하여 위암예측 프로그램을 신규 개발하였다. 또한, 국민건강보험공단 청구자료와 건강검진 자료를 이용한 대장암 발생현황과 대장암 발생에 영향을 주는 연구의 위험요인들을 확인하였다.<sup>22</sup> 폐암은 현재 국립암센터에서 진행 중인 폐암 예측 프로그램과 공동협조로 전수조사를 통하여 진행하였다.

#### 2. 당뇨 질환 예측에 따른 처방메세지 개발

2016년 3차년도 연구에서 개발한 당뇨질환 예측 모형에 대한 건강iN의 마이헬스뱅크에 신규 질환 탑재를 위한 알고리즘 개발 중에 있다.

#### 3. 심뇌혈관 질환 예측 모델 고도화

국민 건강 보험 공단의 건강 검진 자료를 이용하여 비혈액검사적 방법을 통한 심뇌혈관 통합 예측 모델을 개발하였다.

#### 4. 각 생체지표의 연령별, 성별 국가 현황과 생활습관 항목과의 연관성 조사 및 건강노화나이 고도화

기존에 개발한 건강노화나이 모델을 사용한 건강노화나이 등급화를 개발했다. 또한 건강노화나이에 사용되는 각 생체지표의 연령별, 성별 국가 현황 및 흡연, 음주, 운동 등의 생활습관과의 연관성 조사하였다. 이를 사용하여 수검자의 건강노화나이 상태를 제시해 줌으로써 본인의 생활습관 등 건강관리를 집중화 시킬 수 있는 방안을 제시하고자 한다.

## 제2절 연구 방법

### 1. 자료 개요

#### 1) 표본 코호트 DB

건강검진 코호트의 경우 40세 이상의 대상자를 조사한 결과이기 때문에 20~30대의 결과를 반영하기 위해 표본코호트를 활용하였다.

① 기준 : 2002년 자격 대상자(약 100만명)

② 연도 : 2002~2010년(9개년)

〈(참고) 구축 기준 : 2002년 자격 대상자(약 100만명)〉

◆ 표본 추출 : 총 1,476개층(2\*18\*41)

□ 성별(2) : 남자, 여자

□ 연령별(18) : 0세, 1~79세(5세 단위), 80+

□ 소득수준별(41) : 지역 및 직장가입자 각 20분위, 의료급여 0분위

◆ 코호트 구축

□ 연도별 자격 상실자(사망 등)는 코호트에서 자연 탈락

□ 연도별 모집단 신생아 대표 표본은 신규 추가

③내용 : 사회경제적 자격 변수(장애 및 사망 포함), 의료이용현황, 요양기관 현황

<표 3-1> 표본코호트DB 세부특성

구분	세부 특성
자격DB	<ul style="list-style-type: none"> <li>• 대상 : 건강보험가입자 및 의료급여수급권자(외국인 제외)</li> <li>• 변수 : 성, 연령대, 지역, 가입자 구분, 소득분위 등 대상자의 사회경제적 변수 및 장애, 사망관련 총 14개 변수로 구성</li> <li>• 추가자료(2종) : 통계청 사망원인(세분류, 중분류), 시군구자료 → 필요시 검토 후 제공, 통계청 사망자료는 통계청에 제공 내역 통보</li> </ul>
진료DB	<ul style="list-style-type: none"> <li>• 내용 : 대상자가 요양기관에 방문하여 진료 등을 받은 내역에 대해 요양기관으로부터 요양급여가 청구된 자료</li> <li>• 구성 : 의과_보건의료(T1), 치과_한방(T2), 약국(T3)자료에 대한 명세서(20t), 진료내역(30t), 상병내역(40t), 처방전교부상세내역(60t)의 10개 세부DB로 구성</li> <li>• 변수 : 명세서 공통, 진료, 상병, 처방 관련 총 56개 변수로 구성 … (20t) 27개, (30t) 13개, (40t) 5개, (60t) 11개 변수</li> <li>• 의료급여수급권자 2008년 이전 진료상세자료 없음</li> </ul>
건강검진DB	<ul style="list-style-type: none"> <li>• 내용 : 건강검진 주요 결과 및 문진에 의한 생활습관 및 행태관련 자료 … 1차 일반건강검진 자료, 2008년부터 생애전환기건강진단 자료 포함</li> <li>• 구성 : 2002~2008년, 2009~2010년 건강검진DB 별도 구성 … 검진제도 개편(2009년)으로 주요 검진 및 문진항목 변경</li> <li>• 변수 : (2002~2008) 37개 변수, (2009~2010) 41개 변수로 구성</li> </ul>
요양기관DB	<ul style="list-style-type: none"> <li>• 내용 : 요양기관의 종별, 설립구분별, 지역(시도)별 현황 및 시설, 장비, 인력관련 자료</li> <li>• 변수 : 총 10개 변수로 구성</li> </ul>

## 2) 건강검진 코호트 DB

건강검진 수검자 중심의 의료이용 및 건강결과 분석을 위해 2002~2003년 일반 건강검진 수검자 중 2002.12월말 기준 40세~79세의 건강보험 자격유지자 515만명의 10%인 약 51만명에 대한 2002~2013(12년간)의 자격 및 소득정보(사회경제적 변수), 병의원 이용 내역 및 건강검진결과, 요양기관 정보를 코호트형식으로 구축한 개인식별이 불가능한 연구용 DB이다.

## 3) 건강검진코호트DB 표본설계

2002~2003년 일반건강검진 수검자 중 2002.12월말 기준 40세~79세의 건강보험 자격유지자 515만명 중 10% 단순무작위추출을 통해 2002년 자격유지자 중 514,868명을 추출하였다. 이 대상자들을 2013년까지 추적 관찰한 코호트형식의 자료이다.

※ 2002년 의료급여수급자는 일반건강검진 수검 대상이 아니었으므로(2012년 부터 공단에 위탁) 2002년 대상자에서 제외, 이후 연도에는 자격 변동에 따라 가입자 구분이 달라질 수 있다.

※ 만40세미만 수검자는 지역 세대주 및 직장가입자만 검진대상자가 될 수 있다.  
전 국민이 건강검진 대상자가 될 수 있는 시기인 만40세 수검자를 기준으로 표본을 추출한다.

<표 3-2> 건강검진 코호트 DB 보유 정보 주요 내용

구분	세부 특성
자격DB	<ul style="list-style-type: none"> <li>○ 건강보험가입자 및 의료급여수급권자의 건강보장 자격 정보(외국인 제외)               <ul style="list-style-type: none"> <li>- 인구사회학적 정보(성, 연령, 거주지역)</li> <li>- 사망관련 정보(사망일자, 사망원인)</li> <li>- 건강보장 유형(건강보험 가입자 구분/의료급여)</li> <li>- 사회경제적 수준 및 기타 정보 (소득분위, 장애등록정보)                   <ul style="list-style-type: none"> <li>※ 사망 관련 정보의 경우 외부(통계청) 연계 자료 제공</li> </ul> </li> </ul> </li> </ul>
진료DB	<ul style="list-style-type: none"> <li>○ 요양급여비용명세서 상의 의료이용 내역 및 의료비 발생 내역               <ul style="list-style-type: none"> <li>※ 대상자가 요양기관에 방문, 진료 받은 내역에 대하여 요양기관이 청구하여 심사가 결정된 요양급여 내역 자료</li> <li>- 의료기관 이용 정보</li> <li>- 요양급여비용(심사평가원 심사결정요양급여비용)</li> <li>- 진료과목 및 진료 상병 정보</li> <li>- 진찰, 처치, 수술 및 기타 행위 급여 내역, 치료 재료 등 (행위/수가코드) 상세 진료내역</li> <li>- 원내/원외 처방전 의약품 처방내역                   <ul style="list-style-type: none"> <li>※ 요양급여비용명세서 서식 및 유형에 따라 10개 세부DB로 구성</li> <li>※ 의료급여수급권자의 경우, 2006년 이전 진료(청구분) 상세자료 없음</li> </ul> </li> </ul> </li> </ul>
건강검진 DB	<ul style="list-style-type: none"> <li>○ 건강검진 주요 결과 및 문진응답 자료               <ul style="list-style-type: none"> <li>- 일반건강검진 1차, 생애전환기건강진단 1차 검진 결과 및 문진, 구강검진 검진 결과 및 문진                   <ul style="list-style-type: none"> <li>· 신체계측, 혈액 검사 등 주요 검사 결과</li> <li>· 과거력, 생활습관에 관한 문진 결과</li> <li>· 평형성, 골밀도 검사, 우울증, 인지기능 검사 결과 등</li> </ul> </li> <li>※ 의료급여수급권자 일반건강검진 자료는 2012년부터 구축</li> </ul> </li> </ul>
요양기관 DB	<ul style="list-style-type: none"> <li>○ 요양기관 정보               <ul style="list-style-type: none"> <li>- 의료이용 요양기관 종별 및 설립구분, 요양기관 지역정보</li> <li>- 병상 수, 의사 수, 장비 보유 등 요양기관 현황 정보</li> <li>※ 대상자가 이용한 의료기관에 대하여 자료 구축</li> </ul> </li> </ul>

4) 전수자료 DB

맞춤형 자료란 국민건강보험공단이 수집, 보유, 관리하는 모든 수검자의 건강보험 및 장기요양보험 자료에서 연구에 맞는 변수를 요청하여 받은 DB이다.

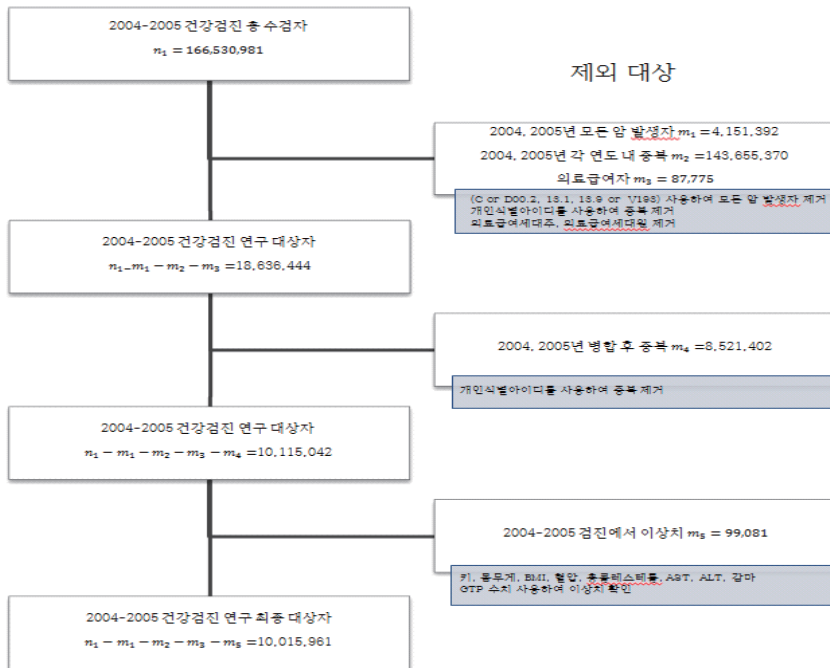
2. 연구대상자

1) 선정 · 제외 기준

표본코호트DB와 건강검진코호트 DB에서 대부분의 대상자들은 2년에 1회 일반건강검진을 시행하므로 2002년도 및 2003년도에 검진 및 문진 내역이 있는 20세 이상의 수검자를 대상으로 한다. 후향적으로 진행되는 탐색적인 연구로 연구대상자 수에 대한 통계적 근거를 두지 않았다.

제3절 각 세부 과제별 연구 방법

1-1. 위암 예측 모형 연구



[그림 3-1] 위암 예측 대상자 선정 순서도

### 1) 대상자

2004~2005년도 국가 일반검진 수검자를 대상으로 1) 상병 코드에 위암 관련 진단코드(C, D00.2, D13.1, D13.9, V193) 기록이 있는 경우, 2) 연구에 사용된 검진자료상 누락된 자료 혹은 극단값(outlier)을 가진 경우, 3) 의료급여자인 경우를 제외하고 2006~2015년 동안 추적하여 구축한 연구용 DB이다. 검진자료상의 극단값은 표본코호트DB의 전체자료를 참고하여 <표 3-3>과 같이 설정하였으며, 총 outlier는 99,081명(0.059 %)이었다.

다음 조건에 따라 최종적으로 모델 구축 및 검정에 포함된 연구 대상자는 10,015,960명이었다. 이 대상자에 대해 최초 건강검진일자부터 2015년까지 진료청구내역서를 기반으로 위암 발생여부를 추적 관찰하였다.

<표 3-3> 위암 예측 모델 구축시 Outlier 제거 기준

	인정 범위
Height	100-220
Weight	30-200
BMI	10-40
BP_Hight	60-240
BP_LWST	40-160
Total Cholesterol	40-400
SGOT_AST	≤200
SGPT_ALT	≤200
Gamma GTP	≤300

### 2) 위암 발생 시간

위암 발생 시간은 처음 검진 받은 날짜를 기준으로 ① 암 발생자의 경우 C16, V193 코드를 받은 날짜, ② 사망자의 경우 사망한 날짜, ③ 암이 발생하지 않은 사람(censored)의 경우 마지막 추적 날짜인 2015년12월31일, ④Lost to follow up(censored)의 경우 마지막 검진 날짜를 사용하였다.

### 3) 위험 요인

모델에 포함된 위험요인으로는 나이, 비만도, 가족력, 식이 습관, 음주량(bottle/day), 흡연량(pack/year), 운동 횟수(per week)를 고려하였다.

#### 4) 모델 개발

모형은 남성과 여성 각각 개발 하였으며, 무작위배정을 통해 7:3으로 모델 구축용(development) 데이터와 검증용(validation) 데이터로 나눈 후 남성과 여성으로 각각 구축하였다. Cox proportional hazards model을 사용하여 단변량, 다변량 분석을 진행하였다. C statistics(area under the ROC curve)을 사용하여 모형의 판별력을 측정하였으며, Calibration plot을 제시하였다.

#### 5) 결측치

남성 위암 모델에서 위험 요인들 중 결측치가 가장 많은 가족력(7.88%)과 흡연량(12.09%)은 결측치를 하나의 범주로 고려하여 대체한 경우와 결측치를 제외한 경우로 2개의 모델을 비교하였다.

### 1-2. 대장암 예측 연구

추후 대장암 예측모델 개발을 위해 대장암 위험요인들을 알아보기 위해 2017년 우리나라 대장암 발병률의 양상과 위험인자에 대한 연구를 참조하였다.<sup>22</sup>

#### 1) 연구자료 및 정의

대장암 발생현황과 대장암 발생에 영향을 주는 위험요인들을 알아보기 위해 국민건강보험공단 청구자료와 건강검진 자료를 이용하였다. 2005년부터 2014년까지 대장암 발생현황을 확인하고, 2004년과 2005년에 건강검진을 받은 사람들을 대상으로 건강검진 요인이 대장암 발생에 미치는 영향을 확인하기 위해 연구 설계를 아래와 같이 진행하였다.

##### ① 대장암 발생위험 인자

최종연구대상자 8,846,749명을 2006년부터 2015년까지 추적관찰하면서 조작적 정의에 해당하는 대장암 대상자의 발생시기와 발생이력을 정리하고 참고문헌들을 통해 정리한 요인들을 모형에 넣어 발생위험 모형을 생성하였다.

##### ② 통계분석방법

새롭게 만들어진 코호트자료의 인구사회학적 특징과 임상적 특징에 따른 대장암발생 위한 분석을 실시하였다. 자료의 분석은 공단자료 접근이 가능한 클라우드 시스템의 SAS Enterprise 통계 프로그램을 사용하였고, 여러

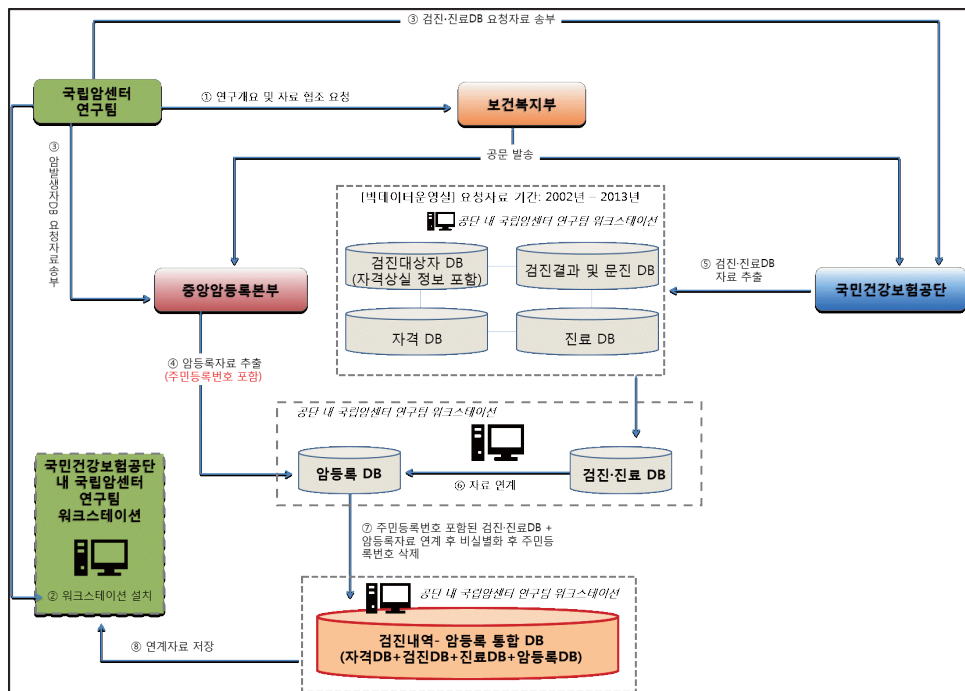


요인들의 기술통계와 빈도분석을 알아보기 위해 chi-square test 와 independent t-test를 실시하였다.

암 경험이 없는 대상자에서 대장암이 발생하기 까지 소요되는 시간과 발생 여부를 이용하여 건강검진정보와 대장암과의 연관성을 알아볼 수 있다. 이에 사용된 분석 모형은 콕스(1972)가 제안한 비례위험모형(Cox-proportional hazard model)을 사용하였다.

### 1-3. 폐암 예측 연구

폐암 검진 대상자 선정을 위한 폐암 발생 위험 예측모델 개발을 위해 자료를 연계하고 모델을 개발하여, 시범사업 기간 내 참여 대상자 선정 시 위험평가에 기반한 예측 모델을 적용해보고자 국민건강보험공단의 검진·진료자료, 국립암센터 중앙암등록본부의 암등록자료를 이용한 폐암 발생 위험 예측 모델 연구 DB를 구축하고, 구축된 DB를 이용하여 모델을 구축 후 개발된 모델의 validation을 시행한다.

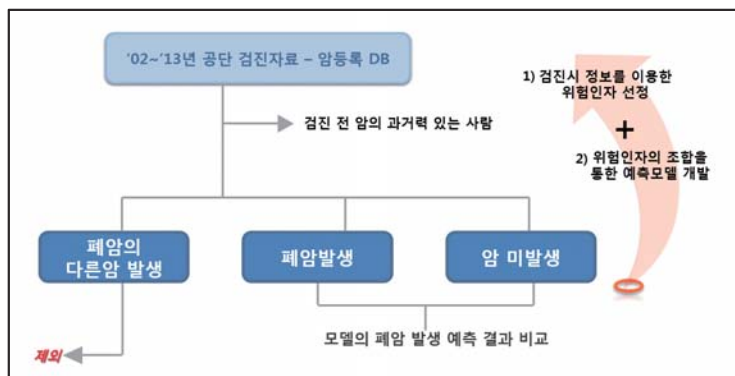


[그림 3-2] 연구 DB 구축 및 자료의 보관·사용 관련 체계

1) 폐암 발생 위험 예측 모델 연구 DB를 구축

기존 연구 데이터베이스가 없기 때문에 국민건강보험공단의 2002년~2013년 까지 일반검진, 암검진을 받은 대상자 전수의 자료를 후향적으로 구득하여, 이 중 암등록자료와 연계 시 검진 전 암으로 진단받았거나 진료 받은 적이 있는 사람, 검진 후 폐암 외에 다른 암이 발생한 대상자는 제외하는 것으로 연구 DB를 설정하였다.

- \* 연구DB: 일반검진, 암검진을 받은 대상자 전수의 자격 DB, 검진 시 문진 및 검진결과를 포함한 검진 DB, 폐암 발생과 관련이 있는 것으로 알려진 일부 질환에 대한 진료 DB 구득하여 구축
- \* 진료 DB: 주상병 및 부상병이 C00~D48, J60, J61, J62, J63, J64, J65, J81, J84 로 기록되어 있는 외래, 입원, 처방기록을 포함 국민건강보험공단에서 구득한 자료에 2014년 암등록자료를 병합(merge)하여 ‘검진 전 암 발생 여부’ 와 ‘검진 후 암 발생 여부’ 확인하고, 이 중 검진 전 암등록자료 상 암 발생이 확인되었거나 암으로 진료받은 기록이 진료 DB에 있는 경우 대상자에서 제외하였다.
- \* 국민건강보험공단 자료와 암등록자료는 암등록자료의 주민등록번호를 통해 자료를 연계(암관리법 시행령 제 24조 2 민감정보 및 고유식별정보의 처리 제①항에 따라 주민등록번호 또는 외국인등록번호가 포함된 자료를 처리할 수 있음)하고 난 후 병합된 자료에 국민건강보험공단의 고유 ID를 부여한 후(비식별화) 개인정보파일은 즉시 파기 처리하였다.



[그림 3-3] 연구 대상자 선정을 위한 연구흐름도

국민건강보험공단 자격, 검진, 진료 DB 내 모델 개발 구축 가능 연도 문진표를 설정하기 위해 흡연 갑년(pack-year)이 산출 가능한 DB로 정하였으나 연도에 따라 흡연 변수에 대한 문진내용이 상이하며, 일부 연도에서만 pack-year가 계산 가능하였다. 폐암 검진 권고안에 따른 폐암 검진 권고 대상자는 현재흡연자 뿐만 아니라 금연한지 15년 이내인 금연자를 포함하고 있기 때문에 금연기간에 대한 정보는 폐암 발생 위험 예측 모델 개발에서 주요 변수이다.

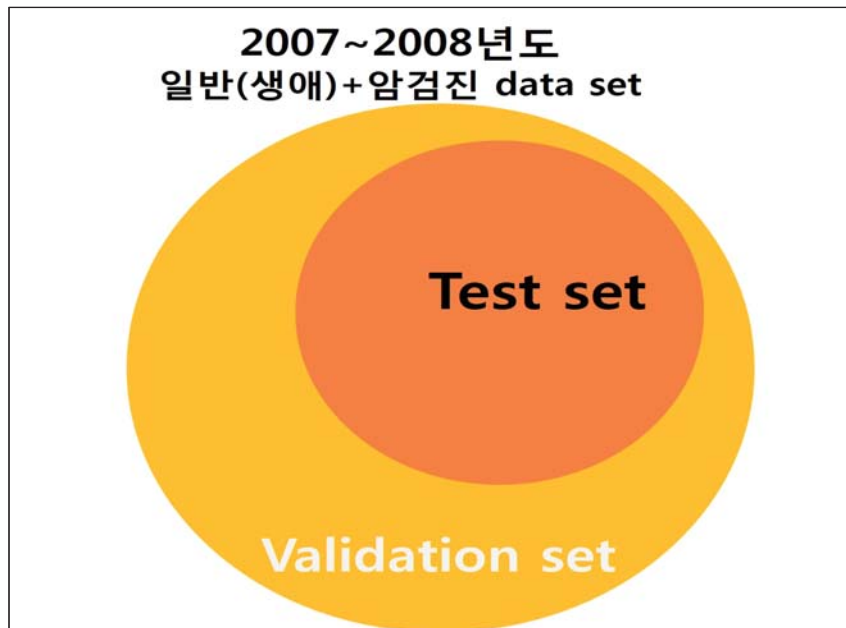
모델 구축에 사용되는 변수 중 2002~2013년까지 국민건강보험공단 연도별 수검자 중 흡연갑년(pack-year)이 과거흡연자도 계산이 가능한 데이터를 모델 구축 데이터로 사용하기로 정하였으며, 현재흡연자 및 과거흡연자의 흡연 갑년(pack-year)이 산출 가능한 DB는 아래와 같다.

<표 3-4> 연도별 문진에 따라 산출 가능한 흡연 변수 항목

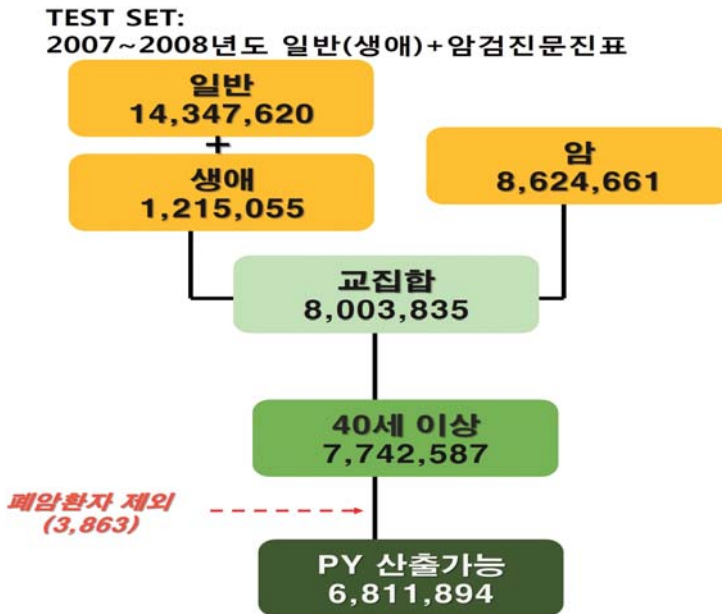
문진대상자		흡연관련	
연도	구분	흡연여부	Pack-year
2002	일반 1차	○	
	생애		
	암	○	
2003	일반 1차	○	
	생애		
	암	○	
2004	일반 1차	○	
	생애		
	암	○	
2005	일반 1차	○	
	생애		
	암	○	
2006	일반 1차	○	
	생애		
	암	○	○
2007	일반 1차	○	
	생애	○	○
	암	○	○
2008	일반 1차	○	
	생애	○	○
	암	○	○
2009	일반 1차	○	○
2010	일반 1차	○	○
2011	일반 1차	○	○
2012	일반 1차	○	○
2013	일반 1차	○	○

- \* 비흡연, 과거 흡연, 현재 흡연 여부: 02-13년 일반 1차 문진, 02-08년 암 문진, 07-08년 생애 전환기 문진
- \* 비흡연, 과거 흡연, 현재 흡연자의 pack-year 범주: 09-13년 일반 1차 문진, 06-08년 암 문진, 07-08년 생애 전환기 문진
- \* 비흡연, 과거 흡연자와 pack-year, 현재 흡연자와 pack-year: 09-13년의 일반 1차 문진, 06-08년의 암 문진, 07-08년의 생애 전환기 문진

정확한 pack-year가 산출 가능한 '07~ '08년의 암 문진, 생애 문진 데이터를 이용하여 2007년 또는 2008년에 일반검진 및 생애검진과 암검진을 모두 받은 수검자 중에 흡연 갑년(PY)이 산출 가능하고 폐암 환자가 아닌 40세 이상 남녀를 연구 대상으로 선정하였고 해당 0708년 연구 대상자는 6,811,894명이었다. 2007~2008년도 일반검진+암검진 data set 중 일부를 test set 설정하고, 그 외 나머지 data를 이용해 validation 진행하였다.



[그림 3-4] 폐암 발생 위험 예측 모델 개발 set 모형도



[그림 3-5] 폐암 발생 위험 예측 모델 개발 set 대상자 수

## 2) 구축된 DB를 이용한 모델 구축 방법

기존에 개발되었던 질병 발생 예측 모델들은 다양한 이론적/통계학적 가설 및 배경을 가지고 있다. 구축된 DB를 이용하여 모델링하기 위해 다음의 방법 중 데이터의 구조와 분포를 확인하여 모델을 구축하기에 적절한 방법을 이용하고자 한다.

### ① 폐암 발생 예측 모델에 포함할 후보 위험인자

검진 DB의 문진, 검진 항목 및 기존에 폐암 발생에 연관성이 있는 것으로 알려진 과거력(기왕 질병)을 문진항목 및 진료 DB의 주상병/ 부상병에서 추출하고, 가용한 후보 위험인자는 다음과 같다.

건강검진 DB 항목(2002.1.1.-2013.12.31.) : 주민등록번호(자료 연계 key)가 포함된 자격DB, 진료DB(의과\_보건기관(T1)자료에서 해당되는 질병코드의 명세서(20t), 진료내역(30t), 상병내역(40t)), 건강검진DB(일반건강검진(`02~`08년), 일반건강검진(`09~`13년), 국가암검진(`02~`13년)

\* 진료자료 DB의 주상병/부상병 항목 : C00-C48(Ⅱ. 신생물, Neoplasms), J60(탄광부진폐증, Coalworker's pneumoconiosis), J61(석면 및 기타

광섬유에 의한 진폐증, Pneumoconiosis due to asbestos and other mineral fibers), J62(실리카를 함유한 먼지에 의한 진폐증, Pneumoconiosis due to dust containing silica), J63(기타 무기물먼지에 의한 진폐증, Pneumoconiosis due to other inorganic dusts), J64(상세불명의 진폐증, Unspecified pneumoconiosis), J65(결핵과 연과된 진폐증, Pneumoconiosis associated with tuberculosis), J81(폐부종, Pulmonary oedema), J84(기타 간질성 폐질환, Other interstitial pulmonary diseases)

② 폐암 발생에 영향을 미치는 인자 선정 방법

(가) Gail 등이 이용한 방법

Gail model은 위험요소의 조합에 따른 상대위험도를 구하고 이를 연령별 유방암 발생률과 유방암을 제외한 연령별 사망률과 합하여 개인의 암 발생의 competing risk(시간에 따른 개인의 암 발생 가능성)을 구한다. 따라서 이 모델을 수행하기 위해 필요한 정보로는  $h1^*$ 는 국립암센터에서 발간하는 ‘국가암등록사업 연례보고서’에서,  $pi(t)$ 는 10년 동안 폐암이 발생한 환자와 발생하지 않은 사람을 환자-대조군 형태의 데이터로 변형하여 각 위험요소 별 대조군의 노출 분율로  $ri(t)$ 는 환자-대조군 형태의 데이터에서 계산한 Odds ratio로,  $h2(t)$ 는 통계청에서 발간하는 한국인 사망원인 중 연령별 사망률과 유방암으로 인한 사망률의 차로 구할 수 있다.

(나) Hazard ratio를 이용한 시간에 따른 절대 위험 상승 정도 관찰

단변량 콕스비례회귀분석을 시행하여 이 중 통계학적으로 유의한 결과를 보인 변수들을 모아 변수선택법을 이용한 다변량 콕스비례회귀분석을 시행하여 한국인에서의 폐암 발생에 관련된 주요 위험요인을 선정한다. 연령별 기저 사망위험도(baseline hazard)를 계산하는데, 이는 연령 특이 사망률에 [각 위험요소의 조합에 따른 사망군의 분포 / 위험도가 가장 낮은 군과 비교하였을 때 각 위험요소의 조합에 따른 사망의 상대위험도]의 합을 곱하여 계산한다. 연령 특이 사망률은 통계청에서 발간하는 사망통계를 이용한다. 기저위험도, 위험요소의 조합에 따른 hazard ratio, 특

정 연령에 도달하기 이전의 사망률을 고려하여 일반인에서 시간에 따른 사망위험도를 계산한다.

③ 모델의 예측력 - area under the receiver operating curve, Hosmer-Lemeshow test를 통해 결정한다.

### 3) 모델의 validation

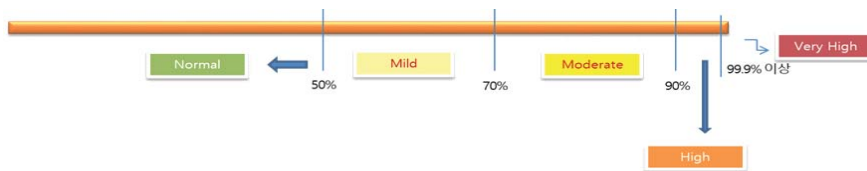
외부 자료를 사용하는 external validation은 현실적으로 자료를 구하기 어렵기 때문에 본 자료를 이용한 internal validation을 시행하게 하고자 한다. 2002년 ~ 2013년까지 검진을 받은 사람은 그 숫자가 충분히 많을 것으로 기대되어 전체 대상자의 30%를 연령, 성별에 따라 층화하여 무작위 추출하여 validation set으로 구축한다. Expected N과 observed N의 비를 구하고 validation set 내에서 area under the receiver operating curve, Hosmer-Lemeshow test를 통해 확인한다.

## 2. 당뇨 질환 예측에 따른 처방메세지 개발

### 1) 당뇨 위험 예측 프로그램 알고리즘 개발

개발한 당뇨 위험 예측 모델을 통하여 당뇨 Cox 회귀분석에 의한 당뇨 위험 예측식을 추정하였다. 추정된 예측식을 통하여, 당뇨 위험 예측도 구간을 설정하였다.

- 50% 미만 : 보통
- 50~70% : 약간 높은
- 70~90% : 높은
- 90~99.9% : 위험
- 99.9% 이상 : 매우 위험



[그림3-6] 당뇨 위험 예측도 구간 설정

당뇨 위험 예측식 및 위험도 분류 구간을 이용, 건강iN 웹페이지 알고리즘을 개발하였다. 당뇨 위험 처방 메시지는 전문가 합의를 통하여 도출하였다. 메시지 구성은 당뇨 위험 메시지, 동일 연령 비교 위험 메시지, 당뇨 위험 교정 가능한 요인 제시 메시지, 주의/위험 분류군 요인별 생활습관 교정 메시지로 구성하였다. 메시지 내용은 건강iN 웹페이지 사용자인 일반인을 대상으로 교육자료의 목적으로 이해하기 쉽도록 대한당뇨병학회, 질병관리본부 당뇨병 예방관리 수칙, 보건복지부 당뇨병 생활습관지침, 국가건강정보포털 등을 참고로 하여 작성하였다.

### 3. 심뇌혈관 질환 예측

#### 1) 연구 대상 및 제외 기준

연구 대상은 건강검진 코호트에서 2002년 혹은 2003년 건강검진을 받은 40세에서 79세 사이의 성인 총 513,340명 중 다음과 같은 제외 기준을 거쳐 선정하였다. 제외 기준은 다음과 같다.

- 건강 검진 전 심뇌혈관 질환 (I20-25, I60-64, I69)를 진단 받거나 관상동맥 재관류요법을 시행 받은 수검자
- 검진 전 심방세동(atrial fibrillation) 을 진단 받은 수검자
- 스타틴을 사용 중인 수검자
- 검진 후 3개월 이내에 본 연구에서 정의하는 심뇌혈관질환이 발생한 환자
- 주요 생활 습관 지표인 흡연력, 운동력, 음주력에 응답하지 않은 수검자
- 검사 결과 이상치(outlier) 에 해당하는 수검자
- 체질량 지수가 12 미만이거나 45 kg/m<sup>2</sup>이상인 경우
- 수축기 혈압이 70mmHg 이하이거나 280mmHg 이상인 경우
- 이완기 혈압이 수축기 혈압보다 높은 경우

제외기준에 부합하는 환자를 제외하고 총 446,783명(남: 243,506명, 여: 203,277명)의 자료를 분석에 사용하였다.

#### 2) 심뇌혈관 질환의 정의

심뇌혈관 질환은 검진 시점 이후 발생한 모든 심근경색(myocardial



infarction)과 뇌혈관 사고(stroke)의 합으로 나타내었다. 심근경색의 정의는 진단 코드(I21-23)를 가지고 1일 이상 입원한 경우와 사망자 중 직접사인분류 코드가 심혈관 질환(I20-I25) 이거나 급성 심장마비(I46) 인 경우로 정의하였다. 뇌혈관 사고는 허혈성, 출혈성, 분류되지 않은 뇌혈관 사고의 진단 코드(I60-I64)와 1일 이상의 입원 및 뇌영상 검사(전산화 단층 촬영이나 자기공명검사)이 발생한 경우와 직접사인분류 코드가 뇌혈관 질환 및 그 합병증(I60~I64, I69)인 경우로 정의하였다.

### 3) 위험 요인 측정

심뇌혈관 질환의 비혈액검사적 위험요인으로 나이, 성별, 체질량 지수, 수축기혈압, 혈압약의 복용 여부, 당뇨병 유무, 흡연력, 음주력, 주간 운동 횟수를 선정하였다. 혈압약의 복용 여부는 수검자의 청구 코드를 확인하여 수검 전 혈압약을 90일 이상 처방받은 경우 혈압약을 복용하는 것으로 정의하였다. 당뇨병의 유무는 건강검진 설문에서 과거력에 당뇨병이 있다고 응답하거나 건진 전 진단코드에서 당뇨병(E10-14)이 2회 이상 진단되거나 당뇨약제를 90일 이상 처방받은 경우로 정의하였다. 흡연력은 설문에서 비흡연자(non-smoker), 과거흡연자(ex-smoker) 와 현재흡연자(current smoker)로 분류하여 정의하였다. 음주력은 설문에서 답변한 일일 음주량과 음주일수를 이용하여 주간 음주약을 계산하여 WHO 기준에 따라 무음주자(non-alcohol user), 저위험음주자(low risk alcohol user), 중위험음주자(intermediate risk alcohol user), 고위험음주자(high risk alcohol user)로 분류하였다.

### 4) 통계적 분석 방법

본 연구에서 연속 변수는 평균과 표준 편차, 명목 변수는 빈도와 백분율(%)을 이용하여 표현하였다. 연속 변수는 independent T test, 명목 변수는 Chi square test를 이용하여 비교하였다. 모델은 남자와 여자 대상자를 구분하여 각각 1:1 무작위 추출 방식을 이용하여 개발군과 검증군으로 나눈 후 개발군에서 Cox 비례 위험 모형을 이용하여 모델을 개발하였다. 연속 변수 중 나이와 체질량 지수는 종속 변수인 심뇌혈관 질환의 발생 위험도와 선형관계가 아니므로 보다 적합한 모델의 개발을 위하여 Multivariable

fractional polynomial regression을 이용하여 종속변수와의 관계를 파악하였다. 또 다른 연속 변수인 수축기 혈압의 경우 임상적으로 반복 측정 시 매번 다르게 측정되는 오차가 있을 수 있어 반복 측정의 오차가 모델에 반영되는 것을 최소화하기 위하여 명목 변수로 치환하여 모델에 적용하였다. 모델에 반영할 변수는 BIC criteria에 의하여 선택하였다. 개발된 모델의 성능은 개발군과 검증군에서 각각 calibration과 discrimination을 측정하였으며 각각 calibration plot 과 concordance index를 계산하여 나타내었다. 추가적으로 계산된 위험도를 바탕으로 저(~3%), 중(3~7.5%) 고(7.5%~) 위험도 군으로 검증군을 분류한 후 각 군의 10년 질환 발생률 그래프를 도시하였다. 또한 개발된 모델을 건강보험공단 샘플코호트 내에서 2009년 건강검진수검자를 대상으로 적용한 후 저, 중, 고위험군에 따라 총 콜레스테롤, 고밀도지방산 콜레스테롤, 중성지방 수치를 비교하여 비혈액검사적 모델로 측정된 위험도가 실제 대상자의 혈액 검사 상의 위험요인과 대응하는지를 알아봄으로써 모델의 성능을 간접적으로 알아보았다. 모든 통계 분석은 R version 3.4.1 을 이용하여 분석하였다.

#### 4. 건강노화나이 연구

##### 1) 연구대상자

국민건강보험공단이 수집, 보유, 관리하는 건강보험 및 장기요양보험 자료 중에서 본 연구에서는 2014-2015년도 맞춤형 자료를 분석하였다. 전체 대상자는 약 2천만명으로 연령은, 20세부터 90세 까지였다. 맞춤형 자료의 전체 대상자들에 대하여 건강노화나이의 생활습관 분석을 시행하였다. 생체지표의 연령별 성별 국가현황 분석은 맞춤형 자료의 전체 대상자 가운데 고혈압, 당뇨병, 심장병, 뇌졸중, (악성)종양이 진단된 대상자와 건강검진 결과가 상대적 정상범위를 벗어난 대상자들을 제외한 약 1천명의 자료를 분석하였다<표3-5>[그림 3-7,8].

신장, 체중, 허리둘레, SBP, DBP, 혈색소, 식전혈당, 총콜레스테롤, 중성지방, 고밀도콜레스테롤, 저밀도콜레스테롤, 요단백, 혈청크레아티닌, eGFR, AST, ALT,  $\gamma$ -GTP

[그림 3-7] 2016년 건강노화나이 검사항목

<표 3-5> 건강인 선정을 위한 생체지표 범위

Inclusion criteria		
Lower (<=)	Parameters	Upper (<)
15.00	BMI (kg/m <sup>2</sup> )	30.00
60.0	WC (cm)	105.0
80.00	SBP (mmHg)	160.00
50.00	DBP (mmHg)	100.00
50.00	FBS (mg/dL)	140.00
50.00	TC (mg/dL)	260.00
50.00	TG (mg/dL)	400.00
20.00	HDL-C (mg/dL)	90.00
-	LDL-C (mg/dL)	190.00
10.00	Hgb (g/dL)	18.00
-	U-Prot (negative, trace, + ~++++)	4.00
0.40	AST (IU/L)	2.00
-	ALT (IU/L)	60.00
-	r-GTP (IU/L)	80.00
-	AST (IU/L)	150.00



[그림 3-8] 건강노화나이 자료분석구조

## 2) 국가 적정 표준화 분류

### ① 생체지표의 연령별, 성별 국가현황

연령대 별 생체지표 변화분석건강노화나이는 일반적으로 생체지표의 조합으로 구성되며 본 연구에서 사용된 생체지표 평가항목은 다음과 같다. 상기 생체지표 항목에 대해 성별 별 연령에 따른 변화분석을 도식화하였다.

② 생활습관의 연령별, 성별국가현황

설문조사를 통해 생활습관 요소인 흡연, 음주, 신체활동에 따른 건강노화나이와 실제나이의 차이를 분석하였다.

흡연의 경우 흡연: 비흡연/과거흡연/현재흡연으로 분류하였으며 음주의 경우 유무로 분류하였다. 신체활동(운동)의 경우, 저강도(주 2일 이내)/중강도(주 3-4일)/고강도(주 5일 이상)으로 분류하여 생활습관에 따른 건강노화나이와 실제나이 차이의 국가별 현황을 제시하였다.

③ 건강노화나이 등급화: 전체, 남녀 별 분석

건강노화나이 등급화에 PARC index를 사용하였다. 건강노화나이와 주민등록상 나이 차이를 PARC index 란 상대적 지표로 나타냈으며 다음과 같은 식을 사용하였다.

$$\begin{aligned} & \text{값이 평균보다 클 때, } (값-평균)/(최대값-평균) \\ & \text{값이 평균보다 작을 때, } -(값-평균)/(최소값-평균) \end{aligned}$$

PARC index의 최대값은 -1이고 최소값은 1로 지정하되 -1에 가까울수록 건강노화나이와 주민등록상 나이차이가 적다는 것을 뜻한다.

# 제4장

## 분석결과

제 1 절 암 질환 예측 연구	45
제 2 절 당뇨 질환 예측에 따른 처방메세지 개발	57
제 3 절 심뇌혈관 질환 예측	67
제 4 절 건강노화나이	73

# 제4장 분석결과

## 제1절 암 질환 예측 연구

### 1. 위암 예측 모형 연구

#### 1) 남성 위암 위험 예측 모델

최종적으로 남성 위암 예측 모델에 포함된 변수는(나이-평균나이), (나이-평균나이)<sup>2</sup>, 비만도, 가족력, 식이습관, 음주량(bottle/day), 흡연량(pack/year), 운동 횟수(per week)이다<표 4-1>. 추적기간 10년간 위암 발생을 확인한 결과 남성 5,482,191명중 86,308명(1.57%) 에서 위암 발생이 확인되었으며, 가족력과 흡연량에서 가장 많은 결측치(7.88%)를 확인 할 수 있었다.

<표 4-1> 위암 예측 모델 분석 대상자의 특성 - 남성

	Development set		Validation set	
	Total (n=3,837,914)		Total (n=1,644,278)	
	Freq	Col%	Freq	Col%
<b>Gastric cancer</b>				
No (0)	3,777,644	98.43	1,618,240	98.42
Yes (1)	60,270	1.57	26,038	1.58
<b>BMI</b>				
<18.5 (1)	88,214	2.30	37,716	2.29
18.5-22.9 (2)	1,340,048	34.92	573,884	34.90
23.0-24.9 (3)	1,048,397	27.32	448,968	27.30
25-29.9 (4)	1,255,266	32.71	538,067	32.72
≥30 (5)	105,990	2.76	45,643	2.78
<b>Family history</b>				
No (1)	3,130,096	81.56	1,341,429	81.58
Yes (2)	405,260	10.56	173,304	10.54
Other (9)	302,559	7.88	129,545	7.88
<b>Meal preferences</b>				
Vegetables (1)	609,340	15.88	261,473	15.90
Mixed (2)	2,935,956	76.50	1,257,617	76.48
Meat(3)	250,447	6.53	107,213	6.52
Missing	42,172	1.10	107,213	6.52

	Development set Total (n=3,837,914)		Validation set Total (n=1,644,278)	
	Freq	Col%	Freq	Col%
Alcohol consumption (bottle/day)				
None (1)	1,259,608	32.82	540,056	32.84
<0.1 (2)	710,254	18.51	304,560	18.52
0.1-0.19 (3)	366,977	9.56	156,730	9.53
≥0.2 (4)	1,441,930	37.57	617,572	37.56
Missing	59,146	1.54	25,360	1.54
Smoking amount (pack/year)				
Non-smoker (1)	1,528,442	39.82	656,152	39.91
Ex-smoker (2)	593,676	15.47	253,946	15.44
<10 (3)	486,698	12.68	208,314	12.67
10-14.9 (4)	51,103	1.33	21,995	1.34
≥15 (5)	714,103	18.61	304,781	18.54
Missing	463,893	12.09	199,090	12.11
Physical activity (per week)				
None (1)	1,731,845	45.12	742,616	45.16
1-2 (2)	1,250,000	32.57	535,960	32.60
Moderate to High (3)	788,352	20.54	336,704	20.48
Missing	67,718	1.76	28,998	1.76

① 다변량 분석 - 남성

위암 예측 모델에 포함된 변수들을 기반으로 다변량(Cox proportional hazard regression) 모델을 구축하였으며, 가) 결측치를 하나의 범주로 고려하여 대체한 경우와 나) 결측치를 대체하지 않은 경우로 2개의 모델을 비교하였다(표 4-2).

(가) 결측치를 대체한 경우

비만도(저체중 HR=1.239, 95% CI:1.181-1.3), 가족력(유 HR=1.265, 95% CI:1.234-1.297), 흡연량(≥15이상 HR=1.319, 95% CI:1.292-1.346)에서 위암 위험을 증가 시켰다.

(나) 결측치를 제외한 경우

비만도(저체중 HR=1.248, 95% CI:1.187-1.313), 가족력(유 HR=1.275, 95% CI:1.243-1.309), 흡연량(≥15이상 HR=1.318, 95% CI:1.291-1.346)에서 위암 발생 위험을 증가 시켰다.

<표 4-2> 나이를 보정한 단변량 모델과 다변량 모델 - 남성 위암 모델

	결측치를 대체한 경우						결측치를 제외한 경우					
	Age-adjusted univariable model (n=3,837,914)			Multivariable model (n=3,837,892)			Age-adjusted univariable model (n=3,837,914)			Multivariable model (n=3,272,819)		
	HR	95% CI	P-value	HR	95% CI	P-value	HR	95% CI	P-value	HR	95% CI	P-value
Age												
Mean (SD)	44.52(13.38)						44.52(13.38)					
Age-Mean (Age-Mean) <sup>2</sup>				1.122	1.12-1.124	<.0001				1.122	1.12-1.124	<.0001
				0.999	0.999-0.999	<.0001				0.999	0.999-0.999	<.0001
BMI												
<18.5 (1)	1.231	1.144-1.326	<.0001	1.239	1.181-1.300	<.0001	1.231	1.144-1.326	<.0001	1.248	1.187-1.313	<.0001
18.5-22.9 (2)	ref			ref			ref			ref		
23.0-24.9 (3)	0.892	0.865-0.920	<.0001	0.925	0.906-0.944	<.0001	0.892	0.865-0.920	<.0001	0.928	0.908-0.948	<.0001
25-29.9 (4)	0.912	0.885-0.939	<.0001	0.947	0.928-0.966	<.0001	0.912	0.885-0.939	<.0001	0.948	0.928-0.968	<.0001
≥ 30 (5)	1.068	0.982-1.163	0.1259	1.117	1.057-1.181	<.0001	1.068	0.982-1.163	0.1259	1.114	1.051-1.182	0.0003
Family history												
No (1)	ref			ref			ref			ref		
Yes (2)	1.323	1.291-1.357	<.0001	1.265	1.234-1.297	<.0001	1.325	1.293-1.358	<.0001	1.275	1.243-1.309	<.0001
Meal preferences												
Vegetables (1)	ref			ref			ref			ref		
Mixed (2)	1.020	0.988-1.053	0.216	0.977	0.957-0.997	0.0272	1.020	0.988-1.053	0.216	0.988	0.967-1.010	0.2774
Meat(3)	1.116	1.045-1.191	0.001	0.967	0.925-1.010	0.1304	1.116	1.045-1.191	0.001	0.986	0.941-1.032	0.5444
Alcohol consumption (bottle/day)												
None (1)	ref			ref			ref			ref		
<0.1 (2)	0.954	0.917-0.992	0.0182	0.899	0.876-0.923	<.0001	0.954	0.917-0.992	0.0182	0.903	0.879-0.928	<.0001
0.1-0.19 (3)	1.017	0.973-1.064	0.4556	0.98	0.952-1.009	0.1794	1.017	0.973-1.064	0.4556	0.981	0.951-1.011	0.2129
≥ 0.2 (4)	1.171	1.138-1.205	<.0001	1.067	1.046-1.088	<.0001	1.171	1.138-1.205	<.0001	1.063	1.042-1.085	<.0001
Smoking amount (pack/year)												
Non-smoker (1)	ref			ref			ref			ref		
Ex-smoker (2)	1.174	1.150-1.203	<.0001	1.153	1.126-1.18	<.0001	1.176	1.150-1.203	<.0001	1.145	1.118-1.172	<.0001
<10 (3)	1.010	0.972-1.050	0.6042	1.145	1.101-1.19	<.0001	1.010	0.972-1.050	0.6147	1.143	1.098-1.189	<.0001
10-14.9 (4)	1.111	1.040-1.187	<.0001	1.09	1.019-1.165	0.0116	1.127	1.091-1.165	<.0001	1.11	0.966-1.275	0.1411
≥ 15 (5)	1.429	1.402-1.458	<.0001	1.319	1.292-1.346	<.0001	1.429	1.420-1.458	<.0001	1.318	1.291-1.346	<.0001
Physical activity (per week)												
None (1)	ref			ref			ref			ref		
1-2 (2)	0.926	0.899-0.955	<.0001	0.932	0.913-0.951	<.0001	0.926	0.899-0.955	<.0001	0.937	0.918-0.957	<.0001
Moderate to High (3)	0.947	0.918-0.976	0.0004	0.95	0.931-0.969	<.0001	0.947	0.918-0.976	0.0004	0.953	0.934-0.974	<.0001



〈추가 모형〉

여성 모델과 비교하기 위해서 남성 위암 모델에서 흡연량, 음주량, 운동 횟수의 범주를 여성 위암 모델과 동일하게 설정 후 분석을 진행하였다(결측치는 제외) 〈표 4-3〉.

<표 4-3> 남성, 여성 위암 모델 변수 차이

남성 위암 모델	여성 위암 모델
Alcohol consumption (bottle/day)	Alcohol consumption (bottle/day)
None (1)	None (1)
<0.1 (2)	<0.1 (2)
0.1-0.19 (3)	≥0.1 (3)
≥0.2 (4)	
Smoking amount (pack/year)	Smoking amount (pack/year)
Non-smoker (1)	Non-smoker (1)
Ex-smoker (2)	Ex-smoker (2)
<10 (3)	
10-14.9 (4)	
≥15 (5)	
Physical activity (per week)	Physical activity (per week)
None (1)	None (1)
1-2 (2)	Yes (2)
Moderate to High (3)	

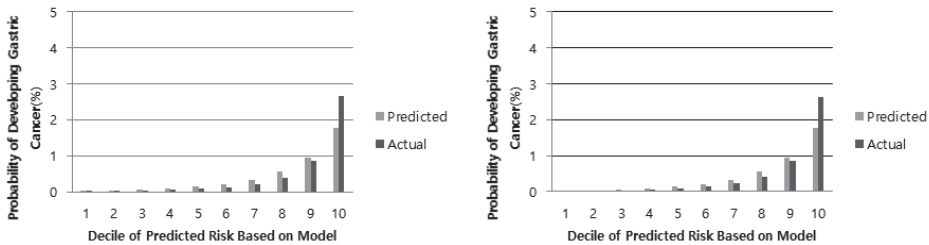
추가 남성 모델의 다변량분석 결과, 나이, 비만도(저체중 HR=1.225, 95% CI:1.192-1.322), 가족력(유 HR=1.274, 95% CI:1.241-1.308), 흡연량 (≥15이상 HR=1.249, 95% CI:1.227-1.271)에서 위암 발생 위험을 증가시켰다 〈표 4-4〉.

<표 4-4> 나이를 보정한 단변량 모델과 다변량 모델 - 남성 추가 모형

	Age-adjusted univariable model (n=3,837,914)			Multivariable model (n=3,047,625)		
	HR	95% CI	P-value	HR	95% CI	P-value
<b>Age</b>						
Mean (SD)	44.52(13.38)					
Age-Mean (Age-Mean) <sup>2</sup>				1.124	1.122-1.126	<.0001
				0.999	0.999-0.999	<.0001
<b>BMI</b>						
<18.5 (1)	1.192	1.137-1.251	<.0001	1.255	1.192-1.322	<.0001
18.5-22.9 (2)	ref			ref		
23.0-24.9 (3)	0.942	0.923-0.962	<.0001	0.922	0.902-0.943	<.0001
25-29.9 (4)	0.973	0.954-0.992	0.0054	0.941	0.921-0.961	<.0001
≥30 (5)	1.119	1.058-1.183	<.0001	1.115	1.049-1.185	0.0004
<b>Family history</b>						
No (1)	ref			ref		
Yes (2)	1.325	1.293-1.358	<.0001	1.274	1.241-1.308	<.0001
<b>Meal preferences</b>						
Vegetables (1)	ref			ref		
Mixed (2)	0.994	0.974-1.015	0.5836	0.995	0.974-1.017	0.6737
Meat(3)	0.99	0.948-1.034	0.6544	1.007	0.96-1.055	0.7841
<b>Alcohol consumption (bottle/day)</b>						
None (1)	ref			ref		
<0.1 (2)	0.932	0.908-0.956	<.0001	0.86	0.838-0.883	<.0001
≥0.1 (3)	1.137	1.117-1.157	<.0001	0.945	0.917-0.974	0.0003
<b>Smoking amount (pack/year)</b>						
Non-smoker (1)	ref			ref		
Current smoker (2)	1.277	1.256-1.298	<.0001	1.249	1.227-1.271	<.0001
<b>Physical activity (per week)</b>						
None (1)	ref			ref		
Yes (2)	0.952	0.936-0.967	<.0001	0.937	0.92-0.953	<.0001

② 남성 모델 평가

남성 위암 모델1에 대한 C statistics는 모델구축용 데이터에서는 0.868[95% CI: 0.866-0.871], 검증용 데이터에서는 0.863[95% CI: 0.859-0.867]로 나타났다. 이는 개발된 모델의 판별력이 상당히 높음을 보여주지만, calibration 결과는 개발된 모델의 예측값과 실제값에 차이가 있음을 나타내었다 (Hosmer-Lemeshow chi-square P<0.0001) [그림 4-1].



[그림 4-1] 남성 위암 모델 calibration 막대 그림(좌: development set, 우: validation set)

2) 여성 위암 위험 예측 모델

최종적으로 여성 위암 예측 모델에 포함된 변수는 (나이-평균나이), (나이-평균나이)<sup>2</sup>, 비만도, 가족력, 식이 습관, 음주량(bottle/day), 흡연량(pack/year), 운동 횟수(per week)이다<표 4-5>. 추적기간 10년간 위암 발생을 확인한 결과 여성 4,443,515명중 35,388명에서 위암 발생이 확인되었으며, 가족력에서 가장 많은 결측치(7.83%)를 확인 할 수 있었다.

<표 4-5> 위암 예측 모델 분석 대상자의 특성 - 여성

	Development set Total (n=3,110,168)		Validation set Total (n=1,333,347)	
	Freq	Col%	Freq	Col%
<b>Gastric cancer</b>				
No (0)	3,086,129	99.23	1,322,969	99.22
Yes (1)	24,039	0.77	10,378	0.78
<b>BMI</b>				
<18.5 (1)	186,284	5.99	79,219	5.94
18.5-22.9 (2)	1,442,027	46.36	620,056	46.50
23.0-24.9 (3)	674,250	21.68	287,591	21.57
25-29.9 (4)	719,825	23.14	308,791	23.16
≥30 (5)	87,782	2.82	37,690	2.83
<b>Family history</b>				
No (1)	2,504,110	80.51	1,072,227	80.42
Yes (2)	362,726	11.66	156,540	11.74
Missing	243,332	7.82	104,580	7.84
<b>Meal preferences</b>				
Vegetables (1)	820,006	26.37	351,328	26.35
Mixed (2)	2,139,171	68.78	917,319	68.80
Meat(3)	112,675	3.62	48,174	3.61
Missing	38,316	1.23	16,526	1.24
<b>Alcohol consumption (bottle/day)</b>				
None (1)	2,268,081	72.92	972,676	72.95
<0.1 (2)	463,379	14.90	198,146	14.86
≥0.1(3)	307,974	9.90	132,303	9.92
Missing	70,734	2.27	30,222	2.27
<b>Smoking amount (pack/year)</b>				
Non-smoker (1)	2,590,895	83.30	1,239,133	92.93
Smoker (2)	205,946	6.62	88,592	6.64
Missing (9)	13,327	0.43	5,622	0.42
<b>Physical activity (per week)</b>				
None (1)	1,904,675	61.24	816,085	61.21
Yes (2)	1,134,920	36.49	487,051	36.53
Missing	70,573	2.27	30,211	2.27

① 다변량 분석 - 여성

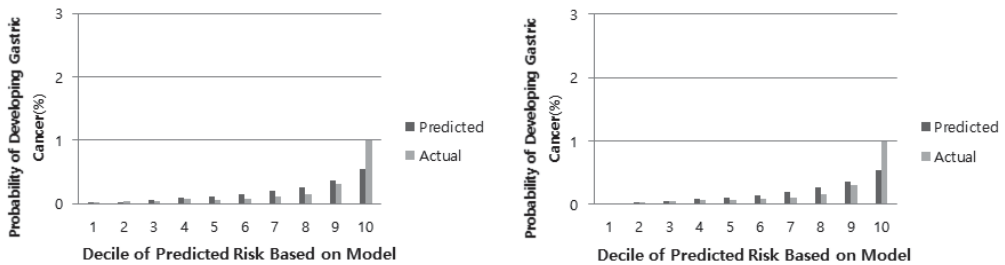
위암 예측 모델에 포함된 변수들을 토대로 다변량 (Cox proportional hazards regression) 모델을 구축하였다. 여성 위암 모델의 다변량 분석 결과 나이, 비만도(저체중 HR=1.209, 95% CI:1.127-1.297), 가족력(유 HR=1.177, 95% CI:1.143-1.211), 흡연량(≥15이상 HR=1.01, 95% CI:0.995-1.099)에서 위암 발생위험을 증가시켰다<표 4-6>.

<표 4-6> 나이를 보정한 단변량 모델과 나이를 보정한 다변량 모델

	Age-adjusted univariable model (n=3,110,168)			Multivariable model (n=3,110,159)		
	HR	95% CI	P-value	HR	95% CI	P-value
<b>Age</b>						
Mean (SD)	46.89 (14.87)					
Age-Mean				1.073	1.071-1.075	<.0001
(Age-Mean) <sup>2</sup>				0.999	0.999-0.999	<.0001
<b>BMI</b>						
<18.5 (1)	1.085	1.011-1.163	0.0227	1.209	1.127-1.297	<.0001
18.5-22.9 (2)	ref			ref		
23.0-24.9 (3)	1.021	0.988-1.055	0.212	0.981	0.95-1.014	0.2664
25-29.9 (4)	0.993	0.962-1.024	0.6492	0.951	0.921-0.981	0.0018
≥30 (5)	1.077	1.007-1.152	0.0315	1.032	0.965-1.104	0.3583
<b>Family history</b>						
No (1)	ref			ref		
Yes (2)	1.311	1.263-1.36	<.0001	1.177	1.143-1.211	<.0001
<b>Meal preferences</b>						
Vegetables (1)	ref			ref		
Mixed (2)	1.002	0.974-1.03	0.899	1.009	0.981-1.037	0.5416
Meat(3)	0.928	0.843-1.022	0.131	0.975	0.885-1.073	0.6024
<b>Alcohol consumption (bottle/day)</b>						
None (1)	ref			ref		
<0.1 (2)	0.952	0.909-0.997	0.0353	0.984	0.94-1.03	0.4892
≥0.1 (3)	1.081	1.026-1.138	0.0033	1.108	1.052-1.166	0.0001
<b>Smoking amount (pack/year)</b>						
Non-smoker (1)	ref			ref		
Current smoker (2)	1.047	0.995-1.101	0.075	1.045	0.995-1.099	0.08
<b>Physical activity (per week)</b>						
None (1)	ref			ref		
Yes (2)	1.058	1.03-1.087	<.0001	1.01	0.985-1.035	0.4352

## ② 모델 평가

여성 모델 평가여성 위암 모델에 대한 C statistics는 development set에서는 0.803[95% CI:0.797-0.81 ], validation set에서는 0.812[95% CI: 0.803-0.821]로 나타났다. 이는 개발된 모델의 판별력이 상당히 높음을 보여주지만, calibration 결과는 개발된 모델의 예측값과 실제값에 차이가 있음을 나타내었다(Hosmer-Lemeshow chi-square  $P < 0.0001$ ) [그림 4-2].



[그림 4-2] 여성 위암 모델 calibration 막대 그림 (좌: development set, 우: validation set)

## 2. 대장암 예측 모형 연구

### 1) 발생의 위험 요인 분석

대장암 발생위험을 조사하기 위해 cox regression 분석을 실시하였고, 각 요인별 통계적 유의성과 발생위험(Hazard Ratio, HR)을 분석하였다. 최종 대장암 발생의 위험요인 모형에 포함된 변수는 연령, 소득수준, 거주지, 몸무게, 공복혈당, 총콜레스테롤, 당뇨 가족력, 암 가족력, 식습관, 음주량, 흡연상태, 운동량 그리고 내시경 검험이 선택되었고, 각 요인의 유의성은 유의수준 0.05 기준으로 하여 분석하였다(표4-7, 8).

<표 4-7> 남성의 대장암 발생 위험요인별 다변량 cox regression 분석결과

		HR	95%CI		p-value
연령	< 40	1			
	40~49	2.685	2.331	3.093	<.0001
	50~59	6.529	5.689	7.492	<.0001
	60~69	13.376	11.636	15.377	<.0001
	70~79	19.177	16.334	22.513	<.0001
	≥ 80	19.19	13.288	27.712	<.0001
소득수준	0~25%				
	25~50%	0.95	0.848	1.066	0.3848
	50~70%	0.955	0.855	1.068	0.42
	70~85%	0.996	0.897	1.107	0.947
	≥ 85%	0.972	0.878	1.076	0.5882
거주지	서울				
	광역시	0.91	0.83	0.997	0.042
	도시	0.981	0.902	1.067	0.6554
	그 외	0.931	0.822	1.054	0.2583
몸무게	< 62				
	62~69	1.071	0.981	1.17	0.125
	69~75	1.111	1.012	1.22	0.0272
	≥ 75	1.254	1.141	1.377	<.0001
공복혈당	< 126				
	≥ 126	1.217	1.092	1.356	0.0004
총콜레스테롤	< 200				
	200~240	0.956	0.891	1.025	0.2012
	≥ 240	1.018	0.92	1.126	0.7343
당뇨가족력	yes	0.91	0.794	1.042	0.1734
암가족력	yes	1.163	1.066	1.269	0.0007
식습관	채식				
	채식, 육식	0.993	0.915	1.077	0.8613
	육식	0.97	0.821	1.145	0.7156
음주량	마시지 않음				
	월 2~3회	1.105	1.007	1.212	0.0349
	주 1~2회	1.217	1.118	1.325	<.0001
	주 3~4회	1.36	1.223	1.511	<.0001
	거의 매일	1.49	1.32	1.682	<.0001
흡연상태	비흡연				
	과거흡연	1.079	0.994	1.173	0.0702
	현재흡연	1.093	1.013	1.18	0.0215
운동량	안한다				
	주 1~2회	0.952	0.882	1.029	0.2175
	주 3~4회	0.994	0.903	1.094	0.8965
	주 5~6회	1.136	0.973	1.328	0.1073
	거의 매일	0.971	0.869	1.085	0.6017
내시경 경험	yes	0.501	0.47	0.533	<.0001

<표 4-8> 여성의 대장암 발생 위험요인별 다변량 cox regression 분석결과

		HR	95%CI		p-value
연령	< 40	1			
	40~49	2.011	1.624	2.49	<.0001
	50~59	2.987	2.407	3.707	<.0001
	60~69	5.207	4.184	6.479	<.0001
	70~79	8.186	6.452	10.386	<.0001
	≥ 80	9.72	5.929	15.934	<.0001
소득수준	0~25%				
	25~50%	1.033	0.901	1.184	0.643
	50~70%	1.058	0.923	1.213	0.4157
	70~85%	0.982	0.86	1.121	0.7864
	≥ 85%	1.169	1.033	1.323	0.0135
거주지	서울				
	광역시	0.71	0.63	0.8	<.0001
	도시	0.889	0.799	0.989	0.0311
	그 외	0.676	0.571	0.799	<.0001
몸무게	< 62				
	62~69	1.11	0.975	1.265	0.1157
	69~75	1.221	1.073	1.388	0.0024
	≥ 75	1.25	1.103	1.417	0.0005
공복혈당	< 126				
	≥ 126	1.183	0.997	1.404	0.0535
총콜레스테롤	< 200				
	200~240	1.067	0.974	1.17	0.1647
	≥ 240	1.059	0.939	1.195	0.3464
당뇨가족력	yes	1.115	0.961	1.294	0.1515
암가족력	yes	1.103	0.99	1.23	0.0758
식습관	채식				
	채식, 육식	1.076	0.983	1.178	0.11
	육식	1.127	0.848	1.497	0.4095
음주량	마시지 않음				
	월 2~3회	0.973	0.844	1.122	0.7053
	주 1~2회	0.925	0.758	1.129	0.4429
	주 3~4회	1.111	0.74	1.668	0.6119
	거의 매일	1.494	0.959	2.327	0.0761
흡연상태	비흡연				
	과거흡연	1.293	0.919	1.821	0.1403
	현재흡연	0.988	0.743	1.313	0.9339
운동량	안한다				
	주 1~2회	1.014	0.912	1.128	0.7945
	주 3~4회	0.902	0.787	1.035	0.1418
	주 5~6회	1.135	0.915	1.407	0.2504
	거의 매일	1.006	0.869	1.164	0.9367
내시경 경험	yes	0.609	0.56	0.663	<.0001

### 3. 폐암 발생 위험 예측

2007, 2008년도 연구 대상자는 6,811,894명 중 pack-year분포는 다음과 같다.

<표 4-9> 폐암 예측 연구대상자 흡연력 분포

	갑년	N	%	총N
비흡연	PY=0	5,306,555	77.90	6,811,894
과거 흡연	0<=PY<10	156,124	2.29	
	10<=PY<20	186,644	2.74	
	20<=PY<30	128,732	1.89	
	30<=PY & 금연>=15Y	19,022	0.28	
	30<=PY & 금연<15Y	134,271		
현재 흡연	0<=PY<10	115,834	1.70	
	10<=PY<20	232,175	3.41	
	20<=PY<30	249,715	3.67	
	30<=PY	282,822	4.15	

#### 1) 2007, 2008년 문진표 내 흡연 외 고려 변수

문진표에 따르면 암 가족력이 매년 증가하며, 실제 암 유병률도 증가한다. 다만, 연도에 따라 암 가족력을 묻는 문항이 다르기 때문에 연도별로 비슷하게 조정하였다. 폐암 발생 위험 예측 모델 개발을 위해 고려하고 있는 폐암 발생 risk factor 중 2007, 2008년도 문진표 내 흡연 외 변수로 사용 가능한 변수는 흡연 시작 연령, 운동량, 시군구(지역), Glucose 레벨 등을 추가 risk factor로 고려하였다.

- ① 음주: 비음주, 과거음주, 현재음주(현재음주자 중 음주량, 빈도도 추후 고려)
- ② 운동: 주당 고강도 운동, 중증도 운동, 걷기 시행 횟수
- ③ 여성: 월경시작나이, 현재월경상태, 폐경연령, 폐경 후 호르몬 대체요법 이용 여부, 출산자녀수, 모유수유여부, 유방양성종양여부, 피임약복용
- ④ 폐결핵: 문진표 상 폐결핵 진단 여부
- ⑤ 기타(검진 결과지): 체질량지수, 혈압(수축,이완기), 요단백, 혈색소수치, 공복혈당수치, 총콜레스테롤수치, SGOT, SGPT, 감마지티피



<표 4-10> 폐암 예측 흡연 외 고려 변수

문진대상자		PY	음주	운동	여성	폐결핵	기타검진
연도	구분						
2007	생애	평생	주간음주일수 일일음주량 최대음주량	고강도 중중도 건기	동일	동일	동일
	암	평생	음주여부 금주후기간 음주횟수 평균음주량	QC_PA_MD_FR Q	동일	동일	동일
2008	생애	평생	주간음주일수 일일음주량 최대음주량	고강도 중중도 건기	동일	동일	동일
	암	평생	음주여부 금주후기간 음주횟수 평균음주량	QC_PA_MD_FR Q	동일	동일	동일

검진 DB 및 암등록 DB를 연계하여 흡연자(비흡연, 과거흡연, 현재흡연) 상황별로 5년 폐암 발생자 수를 산출하였다. 0708년도 연구 대상자는 6,811,894명 중 흡연 갑년별 ‘1,000,000-5 Person Year 폐암 발생자 수’ 는 다음과 같다.

\* 과거흡연자의 경우 금연기간 15년 미만 또는 15년 이상을 구별함

<표 4-11> 1,000,000-5 Person Year 폐암 발생률

	갑년	N	%	5년 폐암 발생자 수	10만명당 폐암발생률	합N	합%	총N
비 흡 연	PY=0	5,306,555	77.90	19,961	376.2	5,306,555	77.90	6,811,894
과거 흡연	0<=PY<10	156,124	2.29	393	251.7	624,793	9.17	
	10<=PY<20	186,644	2.74	732	392.2			
	20<=PY<30	128,732	1.89	867	673.5			
	30<=PY & 금연>=15Y	19,022	0.28	274	1,440.4			
30<=PY & 금연<15Y	134,271	2,312		1,721.9				
현재 흡연	0<=PY<10	115,834	1.70	559	482.6	880,546	12.93	
	10<=PY<20	232,175	3.41	1,691	728.3			
	20<=PY<30	249,715	3.67	2,528	1,012.4			
	30<=PY	282,822	4.15	6,156	2,176.6			

\* 우리나라 2014년 폐암 발생률: 66.0/100,000 Person Year(남) / 28.7/100,000 Person Year(여)

## 제2절 당뇨 질환 예측에 따른 처방메세지 개발

### 1. 당뇨 위험 예측 모델 개발

2016년 당뇨 예측 모델 개발 보고에 따르면, 남자에서 연령, 연령 제곱, 당뇨병 가족력 여부, 비만도, 고혈압 여부, 음주 상태, 흡연 상태, 운동 상태, 총 콜레스테롤, 공복혈당, 이상지질혈증 약 복용 여부로 모형을 개발하였을 때, development set에서의 area under the curve(AUC)는 0.673(95% CI=0.669-0.677)였으며, validation set에서의 AUC는 0.668(95% CI=0.663-0.672)였다. 여자에서 연령, 연령 제곱, 당뇨병 가족력 여부, 비만도, 고혈압 여부, 흡연 상태, 운동 상태, 총 콜레스테롤, 공복혈당, 이상지질혈증 약 복용 여부, GGT로 모형을 개발하였을 때, development set에서의 area under the curve(AUC)는 0.731(95% CI=0.726-0.736)였으며, validation set에서의 AUC는 0.733(95% CI=0.729-0.738)였다 <표4-12>.

<표 4-12> Hazard ratio for related factors for incidence diabetes

	Total			Men			Women		
	HR	(95% CI)		HR	(95% CI)		HR	(95% CI)	
Age, years	1.08	(1.06 - 1.09)	1.06	(1.05 - 1.08)	1.11	(1.09 - 1.14)			
Age squared	1.00	(1.00 - 1.00)	1.00	(1.00 - 1.00)	1.00	(1.00 - 1.00)			
Family history of diabetes	1.32	(1.27 - 1.38)	1.33	(1.26 - 1.41)	1.34	(1.25 - 1.45)			
Body mass index									
<18.5	1.01	(0.92 - 1.09)	1.04	(0.94 - 1.15)	0.92	(0.79 - 1.08)			
18.5-23			Ref.		Ref.				Ref.
23-25	1.27	(1.16 - 1.38)	1.19	(1.08 - 1.32)	1.35	(1.16 - 1.57)			
25-30	1.64	(1.51 - 1.79)	1.48	(1.33 - 1.63)	1.81	(1.56 - 2.11)			
30+	2.47	(2.25 - 2.73)	2.28	(2.01 - 2.58)	2.58	(2.19 - 3.05)			
Hypertension (≥140/90 mmHg and/or using antihypertensive medication)	1.40	(1.37 - 1.44)	1.32	(1.28 - 1.36)	1.45	(1.39 - 1.51)			
Alcohol intake, g/day									
Abstinent			Ref.		Ref.				
Low-medium	1.09	(1.07 - 1.12)	1.03	(1.00 - 1.06)		-			
High	1.30	(1.23 - 1.37)	1.23	(1.16 - 1.30)					
Smoking status									
Non			Ref.		Ref.				Ref.
Former	1.21	(1.16 - 1.25)	1.04	(1.00 - 1.09)	1.40	(1.19 - 1.65)			
Current	1.55	(1.51 - 1.59)	1.33	(1.29 - 1.38)	1.32	(1.20 - 1.45)			
Physically active									
Low			Ref.		Ref.				Ref.
Medium	0.94	(0.92 - 0.97)	0.92	(0.89 - 0.95)	0.92	(0.88 - 0.97)			
High	0.91	(0.88 - 0.95)	0.89	(0.85 - 0.94)	0.93	(0.87 - 0.99)			
Total cholesterol ≥240 mg/dL	1.14	(1.11 - 1.18)	1.13	(1.09 - 1.18)	1.09	(1.04 - 1.14)			
Fasting glucose, mg/dL									
<70	0.99	(0.89 - 1.09)	0.97	(0.86 - 1.10)	0.98	(0.83 - 1.16)			
70-85			Ref.		Ref.				Ref.
85-100	1.22	(1.18 - 1.25)	1.21	(1.17 - 1.26)	1.20	(1.14 - 1.26)			
100-110	2.04	(1.97 - 2.11)	1.93	(1.85 - 2.01)	2.12	(2.00 - 2.24)			
110+	3.49	(3.37 - 3.62)	3.29	(3.14 - 3.43)	3.57	(3.36 - 3.79)			
Using dyslipidemia medication	1.34	(1.28 - 1.41)	1.30	(1.22 - 1.39)	1.33	(1.24 - 1.44)			
GGT, U/L		-		-	1.46	(1.42 - 1.50)			
Development set AUC (95% CI)	0.702	(0.699 - 0.705)	0.673	(0.669 - 0.677)	0.731	(0.726 - 0.736)			
Validation set AUC (95% CI)	0.699	(0.696 - 0.702)	0.668	(0.663 - 0.672)	0.733	(0.729 - 0.738)			

## 2. 당뇨 위험 예측 프로그램 알고리즘 개발

### 1) 당뇨 위험 예측식

개발한 당뇨 예측 모델에 따라서 위험함수를 추정한 당뇨 위험 예측식을 개발하였다.

#### Men

$$\begin{aligned} F(Z) = & 0.0617*(Age-51.1953) - 0.0044*[(Age/10)^2 - 270.5657] + 0.2608*(FamilyhistoryofDM - 0.0584) + 0. \\ & 0261*(BMI < 18.5 - 0.0240) + 0.2074*(BMI_{23-25} - 0.2823) + 0.4276*(BMI_{25-30} - 0.3265) + 0.8712*(BMI \geq \\ & 30 - 0.0195) + 0.2709*(Hypertension - 0.4141) \\ & + 0.1070*(TC \geq 240 - 0.1256) + 0.0050*(Drinkinglow-medium - 0.5910) + 0.1430*(Drinkinghigh - 0.0630) \\ & + 0.0399*(Smokingformer - 0.1563) + 0.2932*(Smokingcurrent - 0.4302) \\ & - 0.0607*(Exercisemedium - 0.4087) - 0.0963*(Exercisehigh - 0.0919) \\ & - 0.0126*(Glucose < 70 - 0.0203) + 0.1948*(Glucose_{85-100} - 0.4158) \\ & + 0.6609*(Glucose_{100-110} - 0.1715) + 1.2066*(Glucose \geq 110 - 0.1040) \\ & + 0.2343*(Using dyslipidemia medication - 0.0420); \\ DM = & EXP(F(Z)), PDM = 1 - S(t)DM = 1 - 0.8995DM \end{aligned}$$

#### Women

$$\begin{aligned} F(Z) = & 0.1010*(Age-52.4922) - 0.0067*[(Age/10)^2 - 284.7600] + 0.3479*(FamilyhistoryofDM - 0.0622) \\ & - 0.0447*(BMI < 18.5 - 0.0234) + 0.2306*(BMI_{23-25} - 0.2623) + 0.5066*(BMI_{25-30} - 0.2941) + 0.8625*( \\ & BMI \geq 30 - 0.0322) + 0.3474*(Hypertension - 0.3371) \\ & + 0.0613*(TC \geq 240 - 0.1514) + 0.2524*(Smokingformer - 0.096) + 0.2140*(Smokingcurrent - 0.0272) \\ & - 0.0714*(Exercisemedium - 0.2432) - 0.0533*(Exercisehigh - 0.0861) \\ & + 0.0192*(Glucose < 70 - 0.0205) + 0.2267*(Glucose_{85-100} - 0.4436) \\ & + 0.7882*(Glucose_{100-110} - 0.1401) + 1.3239*(Glucose \geq 110 - 0.0706) \\ & + 0.3920*(GGT - 2.8405) + 0.2853*(Using dyslipidemia medication - 0.0052); \\ DM = & EXP(F(Z)), PDM = 1 - S(t)DM = 1 - 0.8621DM \end{aligned}$$

## 2) 당뇨 위험 예측도 분류

당뇨 위험 예측식을 이용하여, 10년 당뇨발생 위험예측도를 <표4-13>와 같이 산출하고, 아래와 같이 구간을 분류하였다.

- 50% 미만 : 보통
- 50~70% : 약간 높은
- 70~90% : 높은
- 90~99.9% : 위험
- 99.9% 이상 : 매우 위험

<표 4-13> 10년 당뇨 위험 예측도

10-year probability	Men	Women
50%	0.309	0.129
70%	0.475	0.211
90%	0.762	0.335
99.9%	0.999	0.923

남자의 50%에 해당하는 당뇨 위험 예측도는 30.9%였으며, 여자는 12.9%였으며, 남자 70% 및 여자 70%에 해당하는 예측도는 각각 47.5%, 21.1%였고, 90%에 해당하는 값은 76.2%, 33.5%였다.

## 3) 당뇨 위험 처방 메시지

### ① Message 내용

당뇨 위험 예측 메시지를 아래의 박스와 같이 작성하였다.

현재 생활습관을 10년간 계속 유지할 경우 10년 이내 당뇨 예측 위험도는 보통/약간 높은/높은/위험/매우 위험 수준에 해당하며, 당뇨에 걸릴 확률은 ( )%입니다.

② 위험도 계산에 필요 변수

10년 당뇨 위험 예측도 산출하기 위한 변수는 당뇨위험예측 모델에서 도출한 위험식에 근거하여 아래의 박스와 같다. 변수에 해당하는 저장된 데이터가 있을 시에는 자동으로 데이터를 읽어와 계산하며, 비회원인 경우에는 해당변수를 입력하게 하는 설문을 하여 데이터를 입력하게 하였다.

Men: Age, Age2, Family history of DM, BMI, Hypertension, Alcohol intake, Smoking status, Physical activity, TC, FPG, Using dyslipidemia treatment  
Women: Age, Age2, Family history of DM, BMI, Hypertension, Smoking status, Physical activity, TC, FPG, GGT, Using dyslipidemia treatment

③ 동일 연령 비교 위험 예측 메시지

이용자의 당뇨발생 위험도가 동일 연령에 비교하여 높은지/낮은지 제공하는 메시지를 제공하였으며 효과적인 건강정보 전달 목적으로 작성하였다. 메시지 내용은 아래 박스와 같다.

귀하의 당뇨 예측 위험도는 동일 연령( \_\_세) 평균과 비교하여 낮은/같은/높은 수준에 해당되며, 동일연령 ( \_\_세) 10년 이내 당뇨 예측 위험도 평균은 \_\_% 입니다.

- 낮은 : 동일연령 예측도 > 사용자 당뇨 위험 예측도
- 높은 : 동일연령 예측도 < 사용자 당뇨 위험 예측도
- 같은 : 동일연령 예측도 = 사용자 당뇨 위험 예측도

동일 연령 위험도는 당뇨 위험 예측식에서 연령을 제외한 각 변수들을 평균값으로 변경한 식으로 아래 박스와 같다.

**Men**  
 $F(Z)=0.0617*(Age-51.1953)-0.0044*((Age/10)^2-270.5657);$   
 $DM=EXP(F(Z)), \quad PDM=1-S(t)DM=1-0.8995DM$

**Women**  
 $F(Z)=0.1010*(Age-52.4922)-0.0067*((Age/10)^2-284.7600);$   
 $DM=EXP(F(Z)), \quad PDM=1-S(t)DM=1-0.8621DM$

#### 4) 당뇨 위험도 개선을 위한 생활습관 처방메세지

##### ① 필요한 생활습관 요인

개선이 필요한 생활습관 요인을 제시하기 위한 메시지 내용은 아래 박스와 같다.

개선이 필요한 당신의 생활습관 위험도 개선을 위하여 교정이 필요한 요인은 \_\_\_\_, \_\_\_\_, \_\_\_\_, \_\_\_\_, \_\_가 있습니다.

개선이 필요한 요인은 추정된 당뇨 위험 예측식의 변수들 중에 교정가능한 요인들로 구성되어있다(7가지, BMI, alcohol, smoking, Physical activity, TC, FPG, BP). 교정 가능한 요인(7가지) 중에 “정상” 분류에 해당하지 않는 “주의”, “위험”에 해당하는 요인들을 나열하여 제공하였다. [그림4-3] 과 같이 7가지의 교정 가능한 요인들을 개개인의 데이터 값을 보여주고, 어느 분류에 해당하는지(정상/주의/위험) 제시하며, 권장수치를 제시하였다.

처방메세지 보기
비교그래프 보기

위험요인	귀하의 상태	구분	권장수치(예)
흡연	과거흡연	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #333; color: white; padding: 2px 5px;">정상</div> <div style="background-color: #ccc; padding: 2px 5px;">주의</div> <div style="background-color: #ccc; padding: 2px 5px;">위험</div> </div>	비흡연
음주	저위험 음주	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #333; color: white; padding: 2px 5px;">정상</div> <div style="background-color: #ccc; padding: 2px 5px;">주의</div> <div style="background-color: #ccc; padding: 2px 5px;">위험</div> </div>	절주
운동	충분근	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #333; color: white; padding: 2px 5px;">정상</div> <div style="background-color: #ccc; padding: 2px 5px;">주의</div> <div style="background-color: #ccc; padding: 2px 5px;">위험</div> </div>	충분근
체질량지수(BMI)	29.8	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #ccc; padding: 2px 5px;">정상</div> <div style="background-color: #ccc; padding: 2px 5px;">주의</div> <div style="background-color: #333; color: white; padding: 2px 5px;">위험</div> </div>	23 미만
혈압(수축기/확장기)	130/90	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #ccc; padding: 2px 5px;">정상</div> <div style="background-color: #ccc; padding: 2px 5px;">주의</div> <div style="background-color: #333; color: white; padding: 2px 5px;">위험</div> </div>	수축기 120미만 / 확장기 80미만
공복혈당	126	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #ccc; padding: 2px 5px;">정상</div> <div style="background-color: #ccc; padding: 2px 5px;">주의</div> <div style="background-color: #333; color: white; padding: 2px 5px;">위험</div> </div>	126 이하
총콜레스테롤	210	<div style="display: flex; justify-content: space-around; width: 100%;"> <div style="background-color: #ccc; padding: 2px 5px;">정상</div> <div style="background-color: #333; color: white; padding: 2px 5px;">주의</div> <div style="background-color: #ccc; padding: 2px 5px;">위험</div> </div>	200 미만

[그림 4-3] 당뇨 위험도 개선을 위한 교정 가능한 요인

##### ② 가능한 요인(7가지) 분류 설정

당뇨 위험도 개선을 위한 교정 가능한 요인들의 분류를 [정상], [주의], [위험]군으로 분류하였다. 이용자들이 개개인의 요인별 해당하는 구분이 건강 신호등처럼 어디에 해당하는 지 쉽게 파악할 수 있도록 하여, 손쉬운 건강정보 전달하도록 하였다. 분류 설정은 전문가들의 합의를 통하여 도출하였다.

<표 4-14> 당뇨 위험 교정 가능한 요인 분류 설정

교정 가능한 요인	변수 값	분류군
BMI	<18.5	주의
	18.5-23	정상
	23-25	주의
	25-30	위험
	30+	위험
alcohol	Abstinent	정상
	Low-medium	주의
	High	위험
smoking	Non-smoker	정상
	Former smoker	정상
	Current smoker	위험
Physical activity	Low	위험
	Medium	주의
	High	정상
TC	<200	정상
	200-239	주의
	240 이상	위험
FPG	<100	정상
	100-125	주의
	126 이상	위험
BP	sbp<120 dbp<80	정상
	120<=sbp<140 80<=dbp<90	주의
	Sbp >=140 or Dbp<=90	위험

5) 주의/위험 분류군 처방 메시지

당뇨 위험도 개선을 위한 교정 가능한 요인 7가지에 대하여 “주의”, “위험” 군으로 분류된 생활습관 요인에 대하여 이용자들이 생활습관을 개선 목적으로 메시지를 아래 박스와 같이 작성하였다. 메시지 내용은 건강iN 웹페이지 사용자인 일반인을 대상으로 교육자료의 목적으로 이해하기 쉽도록 대한당뇨병학회, 질병관리본부 당뇨병 예방관리 수칙, 보건복지부 당뇨병 생활습관지침, 국가 건강 정보 포털 등을 참고로 하여 작성하였다.



## 1. 운동

신체활동(운동)이 부족합니다.

당뇨를 예방하기 위하여 꾸준한 신체활동을 시작하시기 바랍니다.

- 운동은 칼로리를 소모시켜 식사요법의 효과를 항진시키고 혈당을 직접적으로 떨어뜨리며 장기적으로 당뇨병의 합병증을 예방하고 스트레스를 해소시키고 정신 건강에도 좋습니다.
- 흡연, 고지혈증, 고혈압, 당뇨 등 심근경색증 및 뇌졸중의 위험요인을 많이 가진 성인일수록 적절한 운동계획을 수립하여 본인의 신체와 심장 기능에 무리가 되지 않도록 주의해야 합니다.
- 운동은 당뇨병환자에서 위축되기 쉬운 근육의 향상을 기대할 수 있으며, 운동으로 인한 말초 조직에서의 인슐린 작용의 증가효과로 고혈당이 개선 될 뿐 아니라 인슐린의 요구량이나 경구 혈당 강하제의 요구량을 감소시킬 수 있습니다.
- 정신적인 안정감, 불안감 해소, 숙면은 물론 현대인의 복잡한 사회생활에서 누적되기 쉬운 스트레스의 해소에도 크게 도움이 되며, 생활에 활력을 얻고 기분이 상쾌하여 삶의 질을 높일 수 있습니다.
- 숨이 조금 찰 정도의 운동의 강도로 하루에 30~60분 가량 실시하는 것이 좋습니다. 무리한 운동으로 너무 많이 땀을 흘리는 것은 오히려 몸을 해칠 수 있으므로 산책, 조깅, 맨손체조, 자전거 타기 등의 가벼운 전신 운동이 좋습니다.
- 매일 30분 정도 빠른 속도로 걷거나, 자전거 타기를 30분 정도 하거나 테니스를 30분 정도 치는 것은 바람직합니다.

## 2) 음주

당뇨를 예방하기 위하여 절주하시기 바랍니다.

절주는 전반적인 건강 상태를 증진시킬 수 있는 가장 중요한 건강습관입니다.

- 술은 영양소가 들어 있지 않은 고열량식품이기 때문에 더 많은 인슐린분비를 요구하게 되어 췌장의 베타세포에 많은 부담을 주고, 체내에서 알코올은 지방산의 합성을 증가시켜서 인슐린 저항성을 일으킵니다.
- 과도한 음주는 공복혈당을 상승시키고, 당뇨병의 발생률을 높이고, 지방간 및 간경화 췌장염, 심혈관 질환의 위험을 높입니다.
- 당뇨병 예방을 위하여 남자의 경우 하루 해당 주류 잔으로 2잔 이하, 여성은 1잔 이하로 적정음주를 유지하시기 바랍니다.

- 하루에 한두 잔 이하의 음주는 혈관이 좁아지거나 막혀서 생기는 허혈성 뇌졸중 및 심혈관질환의 예방에 도움이 될 수도 있으나, “중용적 음주” 혹은 “절제적 음주”란 실제적으로 어려운 일이므로 차라리 절주 보다는 금주하는 것이 권장됩니다.

### 3) 총콜레스테롤

귀하의 콜레스테롤은 ( )mg/dL 입니다.

정상 200mg/dL미만 목표로 혈중 콜레스테롤 조절을 시작하여 뇌졸중 위험을 낮추십시오.

- 고지혈증은 글자 그대로 피 속에 기름(지방)이 정상보다 많다는 것입니다. 그 원인은 우리가 알고 있는 것처럼 고기를 많이 먹어서만 오는 것은 아닙니다. 과잉 칼로리 섭취에 의해서, 술이나 약에 의해서, 유전적인 영향에 의해서, 또는 어떤 질환에 의해 이차적으로 올 수도 있습니다.
- 우리 몸속에 콜레스테롤이 필요 이상으로 많아질 경우 지방 성분이 혈관 벽에 쌓여 동맥경화, 고혈압, 뇌졸중 등 심혈관계 질환과 당뇨 등의 성인병을 일으키는 데 큰 영향을 미칩니다.
- 동물성 기름에는 포화지방산이 많아 혈중 콜레스테롤을 높이므로, 기름은 식물성기름으로 사용하고 콜레스테롤이 많은 음식은 1주일에 2~3회 이하로 섭취를 제한합니다.
- 식이요법(삼겹살 등 기름진 음식·피자·계란노른자·명란젓·알탕·곱창·간·새우·오징어·문어·낙지 섭취 줄이기, 잡곡류·채소 섭취 늘리기), 운동요법(조깅, 수영 등 유산소 운동)이 필요합니다.

### 4) 흡연

당뇨를 예방하기 위하여 금연하시기 바랍니다.

금연은 전반적인 건강 상태를 증진시킬 수 있는 가장 중요한 건강습관입니다.

- 흡연은 주요 심뇌혈관질환에 걸릴 위험이 비흡연자들에 비해 2배정도 높습니다. 흡연자에게 최선의 치료는 금연입니다.
- 금연을 하면 암을 비롯한 여러 질환 발생의 위험이 감소되고 건강하고 깨끗한 인상을 만드는데 도움이 됩니다. 그 밖에도 입과 옷 머리 등에 밴 나쁜 냄새가 없어지며, 치아가 하얗고 건강해지며, 손가락의 착색이 사라집니다. 후각이 돌아와 식욕도 좋아지며, 계단을 오를 때

숨이 덜 차게 되는 등 흡연으로 인해 제대로 기능하지 못했던 신체가 회복되므로 꼭 금연하 시기 바랍니다

- 주위의 금연 방해요인들(금단증상, 스트레스, 동료나 주위의 압력, 체중증가, 실패에 대한 두 려움)이 있을 수 있으므로 스스로 금연이 어려울 경우 금연클리닉, 금연콜센터(1544-9030) 와 같은 금연지원 서비스를 이용하는 것이 도움이 됩니다.

## 5) 혈압

귀하의 혈압은 ( )/( ) mmHg입니다.

정상 기준치 120/80mmHg 보다 높으므로 뇌졸중 위험을 낮추기 위하여 혈압 조절을 시작하십 시오.

- 고혈압은 수축기 혈압 140mmHg 이상, 이완기 혈압 90mmHg 이상을 말하며, 진행되기 전 까지는 뚜렷한 증상이 없으며, 심뇌혈관질환의 위험이 증가합니다.
- 염분 섭취를 줄이기 위해 김치, 장아찌, 젓갈, 소시지, 햄, 치즈 등의 섭취를 줄이며, 조리 시 에 된장, 고추장, 간장, 소금 등은 줄여서 사용하고 화학 조미료는 사용하지 않도록 합니다. 카페인 많은 식품(홍차, 커피, 초콜릿, 콜라 등) 섭취는 줄이고, 음주는 적정횟수 및 적정량 을 초과하지 않도록 합니다. 주 3회(1회 30분 이상) 유산소 운동을 유지해 주십시오.
- 정기적인 혈압측정을 하시기 바라며, 합병증을 예방하기 위해 약물치료가 필요할 수 있으므 로 의사와의 상담을 권유드립니다.

## 6) 비만

귀하의 체질량지수는 ( )kg/m<sup>2</sup>으로 정상 23kg/m<sup>2</sup>보다 높습니다.

당뇨 위험을 낮추기 위하여 체중조절을 시작하십시오

- 비만은 당뇨 발생위험을 증가시키며 체중감량을 통해 혈압을 낮추고 당뇨 발생을 예방할 수 있습니다.
- 비만은 당뇨병을 비롯한 고혈압, 이상지질혈증, 뇌졸중, 관상동맥 질환 등과 밀접한 관련이 있습니다. 우리나라에서는 체질량 지수가 25 kg/m<sup>2</sup> 이상이면 비만으로 정의하며, 23 kg/m<sup>2</sup> 이상이라도 과체중이며 당뇨병 위험 요인으로 간주합니다
- 식사 조절과 운동 등으로 체중을 7% 줄이면 당뇨병 발생을 58% 정도 줄일 수 있습니다. 그

뿐만 아니라 혈압도 낮추고, 이상지질혈증도 개선됩니다. 이러한 변화들은 대사 질환과 심혈관 질환을 예방하고, 치료하는 데도 중요합니다. 따라서 당뇨병 발병 위험이 높은 비만인 사람들에게는 체중의 5~10%를 감량하도록 권장합니다.

- 비만치료에서 가장 효과적인 방법은 식사조절과 운동을 병행하는 것으로 효과적인 체중감량을 위해서는 열량 섭취를 제한하여야 하며, 열량 섭취 제한 정도는 개인의 상태를 고려하여 결정해야 합니다.

### 7)혈당

귀하의 공복혈당은 ( )mg/dL 으로 정상 기준치 100mg/dL 이상입니다.

당뇨 위험을 낮추기 위하여 혈당조절을 시작하십시오.

- 공복혈당장애는 8시간 금식 후 측정된 혈당이 정상보다는 높지만 당뇨병 진단 기준보다는 낮은 상태로 혈당 수치 100~126mg/dL 사이에 해당됩니다. 상 혈당인 사람이 당뇨병에 걸릴 위험이 1년에 약 0.7%이지만, 공복혈당장애가 있는 사람은 당뇨병 발생이 정상인에 비하여 10배 가량 높으며, 심혈관계 질환 발생도 정상인에 비하여 2배 가량 높습니다.
- 만성적인 고혈당은 신체 각 기관의 손상과 기능 부전을 초래하게 됩니다. 망막, 신장, 신경에 나타나는 미세혈관 합병증이 있으며, 동맥경화, 심혈관, 뇌혈관질환과 같은 거대 혈관 합병증을 유발하고 이로 인한 사망률을 증가시킵니다.
- 적정 체중 유지, 식습관 조절(당분 및 탄수화물 섭취 줄이기/잡곡, 채소 섭취 늘리기)과 규칙적인 운동(주 3회(1회 30분 이상) 유산소 운동)을 유지하시기 바랍니다.

## 제3절 심뇌혈관 질환 예측

### 1. 심뇌혈관 질환 발생 여부에 따른 연구 대상의 특성

전체 환자 대상군에서 총4,917,150 인-년(person-year)의 추적 기간 동안 남성은 약7.0%인 17165명에서 여성은 약 5.9%인 11238명에서 심뇌혈관질환이 발생하였다. 심뇌혈관 질환의 발생 여부에 따른 특성은 <표 4-15>에 도시하였다.

<표 4-15> 심뇌혈관 질환 발생 여부에 따른 대상군의 특성

Followed ASCVD	Male (N=243,506)		P	Female (N=203,277)		P
	Yes (N=17,165)	No (N=226,341)		Yes (N=11,238)	No (N=192,039)	
Age, years	59.110.0	51.58.9	<0.001	62.39.7	52.89.3	<0.001
40~49	3,604 (21.0%)	113,616 (50.2%)		1,440 (12.8%)	83,971 (43.7%)	
50~59	4,642 (27.0%)	65,291 (28.8%)		2,298 (20.4%)	56,812 (29.6%)	
60~69	5,852 (34.1%)	37,423 (16.5%)		4,379 (39.0%)	38,992 (20.3%)	
70~79	3,067 (17.9%)	10,011 (4.4%)		3,121 (27.8%)	12,264 (6.4%)	
BMI, kg/m <sup>2</sup>	23.73.0	24.02.8	<0.001	24.33.3	23.93.0	<0.001
~18.4	686 (4.0%)	5,137 (2.3%)		366 (3.3%)	4,304 (2.2%)	
18.5~22.9	6,289 (36.6%)	77,840 (34.4%)		3,717 (33.1%)	73,998 (38.5%)	
23~24.9	4,482 (26.1%)	64,249 (28.4%)		2,782 (24.8%)	50,243 (26.2%)	
25~29.9	5,349 (31.2%)	74,526 (32.9%)		3,819 (34.0%)	56,936 (29.6%)	
30~	359 (2.1%)	4,589 (2.0%)		554 (4.9%)	6,558 (3.4%)	
Diabetes mellitus	1,999 (11.6%)	11,621 (5.1%)	<0.001	1,358 (12.1%)	7,936 (4.1%)	<0.001
Ant-hypertensive user	3,482 (20.3%)	23,539 (10.4%)	<0.001	3,607 (32.1%)	30,310 (15.8%)	<0.001
SBP, mmHg	134.819.6	128.017.1	<0.001	133.921.0	124.118.4	<0.001
~119	2,731 (15.9%)	56,959 (25.2%)		2,256 (20.1%)	70,908 (36.9%)	
120~139	7,135 (41.6%)	108,056 (47.7%)		4,366 (38.9%)	78,888 (41.1%)	
140~159	5,124 (29.9%)	48,207 (21.3%)		3,167 (28.2%)	32,432 (16.9%)	
160~	2,175 (12.7%)	13,119 (5.8%)		1,449 (12.9%)	9,811 (5.1%)	
DBP, mmHg	83.512.3	81.011.3	<0.001	81.312.4	77.111.7	<0.001
~69	1,258 (7.3%)	20,641 (9.1%)		1,265 (11.3%)	36,230 (18.9%)	
70~79	3,373 (19.7%)	55,972 (24.7%)		2,617 (23.3%)	57,287 (29.8%)	
80~89	6,011 (35.0%)	83,966 (37.1%)		3,778 (33.6%)	60,617 (31.6%)	
90~	6,523 (38.0%)	65,762 (29.1%)		3,578 (31.8%)	37,905 (19.7%)	
Smoking status			<0.001			<0.001
None	6,917 (40.3%)	94,838 (41.9%)		10,469 (93.2%)	185,082 (96.4%)	
Ex-smoker	2,158 (12.6%)	34,934 (15.4%)		122 (1.1%)	1,809 (0.9%)	
Current smoker	8,090 (47.1%)	96,569 (42.7%)		647 (5.8%)	5,148 (2.7%)	
Alcohol consumption			<0.001			<0.001

Followed ASCVD	Male (N=243,506)		P	Female (N=203,277)		P
	Yes (N=17,165)	No (N=226,341)		Yes (N=11,238)	No (N=192,039)	
None	6,962 (40.6%)	77,265 (34.1%)		9,719 (86.5%)	156,884 (81.7%)	
Low risk	7,721 (45.0%)	124,663 (55.1%)		1,359 (12.1%)	32,333 (16.8%)	
Moderate risk	279 (1.6%)	5,068 (2.2%)		48 (0.4%)	1,106 (0.6%)	
High risk	2,203 (12.8%)	19,345 (8.5%)		112 (1.0%)	1,716 (0.9%)	
Exercise per week			<0.001			<0.001
None	9,909 (57.7%)	110,417 (48.8%)		8,361 (74.4%)	127,369 (66.3%)	
1~2	3,859 (22.5%)	69,134 (30.5%)		1,337 (11.9%)	32,937 (17.2%)	
3~4	1,453 (8.5%)	25,594 (11.3%)		508 (4.5%)	14,598 (7.6%)	
5~6	404 (2.4%)	6,338 (2.8%)		168 (1.5%)	4,385 (2.3%)	
Everyday	1,540 (9.0%)	14,858 (6.6%)		864 (7.7%)	12,750 (6.6%)	

심뇌혈관 발생의 위험요인에 대하여 <표 4-16>에 나타내었다. 남녀 모두 심뇌혈관 질환이 발생한 환자의 경우 고령, 당뇨, 고혈압 약제 사용률, 높은 혈압, 흡연력 등이 위험요인으로 확인되었다. 체질량 지수의 경우 저체중(18.4kg/m<sup>2</sup> 이하) 이거나 비만(25 kg/m<sup>2</sup> 이상)인 경우, 음주력의 경우, 음주는 전혀 하지 않거나 고위험음주를 하는 경우, 운동의 경우 전혀 하지 않거나, 매일 운동을 하는 경우 심뇌혈관계 위험도가 높아지는 양상이었다.

<표 4-16> 심뇌혈관 질환 발생의 위험 요인

	Male			Female		
	HR	95% CI	P	HR	95% CI	P
Age, years	1.085	(1.083, 1.086)	<0.001	1.099	(1.097, 1.101)	<0.001
BMI, kg/m <sup>2</sup>						
~18.4	2.056	(1.897, 2.228)	<0.001	1.702	(1.528, 1.895)	<0.001
18.5~22.9	1.173	(1.129, 1.219)	<0.001	Reference	-	-
23~24.9	Reference	-	-	1.101	(1.048, 1.156)	<0.001
25~29.9	1.024	(0.984, 1.065)	0.250	1.327	(1.269, 1.389)	<0.001
30~	1.113	(1.000, 1.240)	0.050	1.660	(1.518, 1.815)	<0.001
Diabetes mellitus	2.423	(2.312, 2.538)	<0.001	3.121	(2.949, 3.303)	<0.001

Ant-hypertensive user	2.189	(2.109, 2.272)	<0.001	2.508	(2.410, 2.609)	<0.001
SBP, mmHg						
~119	Reference	-	-	Reference	-	-
120~139	1.371	(1.312, 1.433)	<0.001	1.733	(1.647, 1.823)	<0.001
140~159	2.201	(2.101, 2.305)	<0.001	3.029	(2.870, 3.197)	<0.001
160~	3.425	(3.237, 3.623)	<0.001	4.528	(4.239, 4.837)	<0.001
Smoking status						
None	Reference	-	-	Reference	-	-
Ex-smoker	0.845	(0.806, 0.887)	<0.001	1.194	(0.999, 1.427)	0.052
Current smoker	1.147	(1.111, 1.184)	<0.001	2.203	(2.035, 2.385)	<0.001
Alcohol consumption						
None	1.465	(1.418, 1.513)	<0.001	1.464	(1.383, 1.549)	<0.001
Low risk	Reference	-	-	Reference	-	-
Moderate risk	0.888	(0.788, 1.001)	0.052	1.038	(0.778, 1.384)	0.801
High risk	1.864	(1.778, 1.954)	<0.001	1.549	(1.278, 1.878)	<0.001
Exercise per week						
None	1.595	(1.509, 1.685)	<0.001	1.870	(1.710, 2.045)	<0.001
1~2	0.985	(0.928, 1.046)	0.626	1.165	(1.052, 1.290)	0.003
3~4	Reference	-	-	Reference	-	-
5~6	1.117	(1.001, 1.247)	0.049	1.102	(0.926, 1.312)	0.275
Everyday	1.831	(1.704, 1.967)	<0.001	1.931	(1.730, 2.154)	<0.001

## 2. 새로운 심뇌혈관 통합예측모델 및 성능 평가

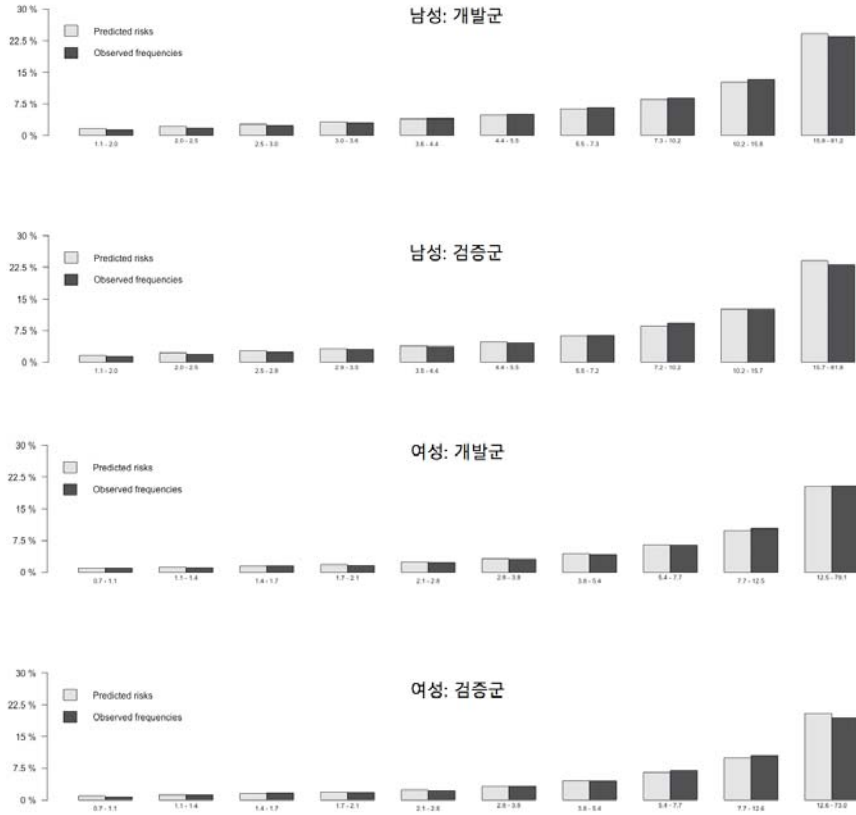
비혈액검사적 요인을 사용하여 심뇌혈관 통합 예측 모델은 <표4-17>에 도시하였다. 남성 모델의 경우 나이와 수축기 혈압, 고혈압 약제 복용 여부, 당뇨, 체질량 지수, 흡연력, 운동습관 및 음주 습관이 최종 모델의 변수에 포함되었으나 여성의 경우에는 음주 습관은 최종 모델에서 배제되었다. 이는 여성의 경우 전체 응답자의 80% 이상이 음주력이 없는 것으로 응답하였으며 고위험 음주군이 연구 대상의 약 1% 정도에 불과하여 모델의 성능을 높이는 변별력을 갖기 어려웠던 것으로 파악된다. 최종 예측 모델의 경우 남성 개발군에서 0.760, 검증군에서 0.752의 C-index를 보였으며 여성의 경우 개발군에서 0.792 검증군에서 0.788의 C-index를 보였다.

<표 4-17> 남녀에서의 새로운 심뇌혈관 통합 예측 모델

	Male				Female			
	Beta	SE	HR	P	Beta	SE	HR	P
Age	0.07506	0.00119	1.078	<0.001	0.0833447	0.0014880	1.087	<0.001
SBP: 120~139 mmHg	0.2239	0.03182	1.251	<0.001	0.2257204	0.0374125	1.253	<0.001
SBP: 140~159 mmHg	0.4330	0.03449	1.542	<0.001	0.3639237	0.0414635	1.439	<0.001
SBP: 160~ mmHg	0.6713	0.04241	1.957	<0.001	0.5514701	0.0508778	1.736	<0.001
Anti-hyp ertensive user	0.2325	0.02846	1.262	<0.001	0.3121178	0.0305428	1.366	<0.001
DM	0.5790	0.03396	1.784	<0.001	0.6241892	0.0412837	1.867	<0.001
10/BMI	-1.808	0.4242	0.1640	<0.001	-1.1038536	0.4281873	0.332	0.010
Log(BMI/ 10)*10/B MI	-11.14	1.986	0.00001459	<0.001	-8.6661156	2.1913521	0.0001723	<0.001
Current smoker	0.4335	0.02263	1.543	<0.001	0.5533200	0.0577688	1.7390170	<0.001
Exercise: none	0.1718	0.02235	1.187	<0.001	0.1716138	0.0357629	1.1872193	<0.001
Exercise: everyday	-	-	-	-	0.1393770	0.0571793	1.1495574	0.014
Alcohol consump tion: none	0.1162	0.02427	1.123	<0.001	-	-	-	-
Alcohol consump tion: high risk	0.1373	0.03470	1.147	<0.001	-	-	-	-
Baseline survival		0.9468336				0.9647013		
C index: develop ment set		0.760				0.792		
C index: validation set		0.752				0.788		

새로운 예측 모델의 성별에 따른 Calibration plot을 [그림 4-4]에 도시하였다. 남성과 여성 모두 고위험군에서 위험도를 과다추정하는 경향이 발견되었으나 실제 값의 차이가 크지 않아 용인할 수 있는 수준의 오차를 보인 것으로 판단되었다.

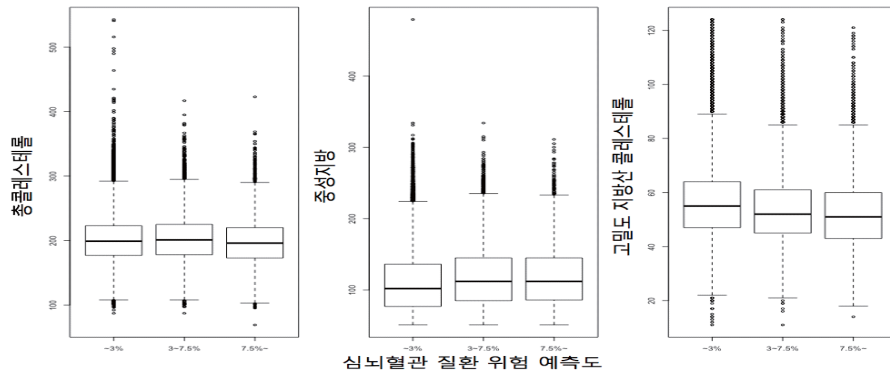




[그림 4-4] 심뇌혈관 통합 예측 모델의 Calibration plot

### 3. 심뇌혈관 질환 위험도에 따른 혈중 콜레스테롤의 변화

새로 개발한 비혈액검사적 심뇌혈관 질환 예측 모델이 실제 혈액 검사 상 알려진 대표적인 심뇌혈관 질환의 위험요인인 혈중 콜레스테롤 수치와 대응하는지 여부를 확인하여 예측 모델이 실제 합리적인 추정을 하는지 여부를 확인하기 위하여 심뇌혈관질환의 위험도에 따른 혈중 콜레스테롤 수치를 확인하여 [그림 4-5] 에 나타내었다. 심뇌혈관 질환의 위험도가 높을수록 중성지방이 높고 고밀도지방산 콜레스테롤이 낮아지는 것을 관찰할 수 있었다. 이는 좋지 않은 생활습관이 고위험의 혈중 콜레스테롤 수치와 연관이 있음을 보여주었다. 총콜레스테롤 수치는 예측된 심뇌혈관 질환의 위험도에 따른 차이를 보이지 않았다.



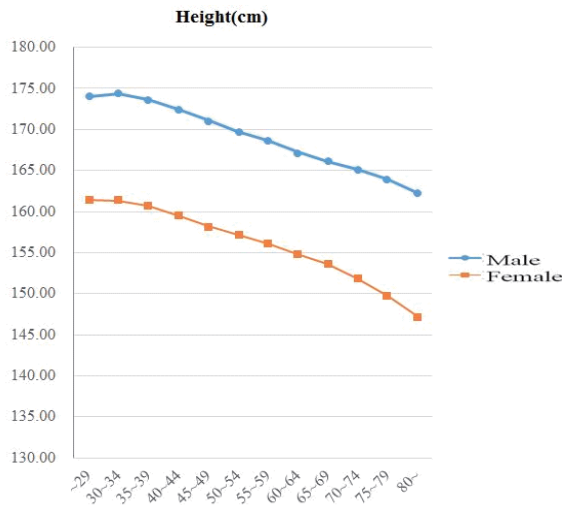
[그림 4-5] 10년 심뇌혈관 발생 위험도에 따른 혈중 콜레스테롤 수치의 변화

## 제4절 건강노화나이

### 1. 연령에 따른 생체지표 변화

#### 1) Height

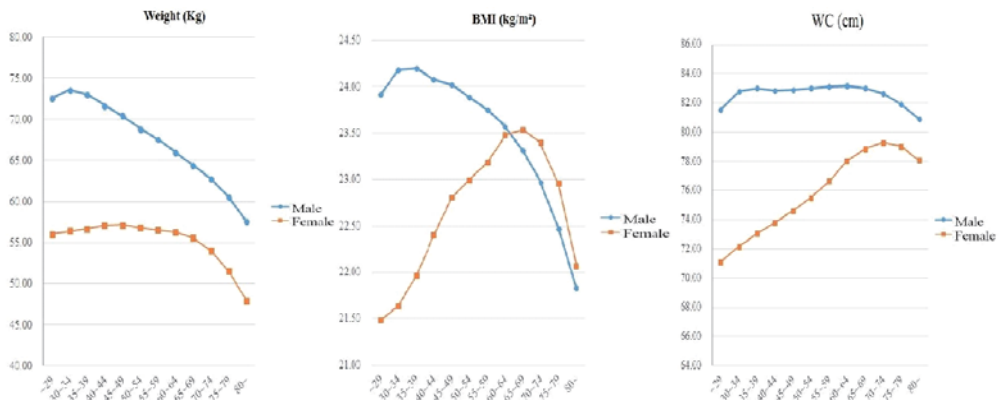
연령증가에 따른 신장의 변화는 [그림4-6] 과 같이 남녀 모두 감소하는 경향을 보였다. 기존연구에 따르면 연령 증가로 인한 신장의 감소는 45세 정도에서 시작되며 이는 주로 척추 골절, 자세변화 등의 척추의 변화에 기인하는 것으로 알려져 있다.<sup>23</sup>



[그림 4-6] 연령에 따른 신장 변화

## 2) Weight, BMI, WC

체중의 변화는 [그림4-7]과 같이 연령 증가함에 따라 남녀 모두에서 감소하는 경향을 보이며 특히 여자보다 남자에서 감소하는 경향이 더 뚜렷이 나타난다. 기존연구에서 남성의 체중은 50세까지 증가하다 그 이후는 약간 감소하고 여성의 경우는 60세까지 체중이 전반적으로 증가하며, 그 이후에는 약간 감소하거나 변화가 없는 것으로 알려져 있으며 이 같은 결과는 본 연구 결과와 차이를 보인다.<sup>24</sup> BMI의 경우 남자에서는 20대 이후 감소 추세를 보이거나 여성의 경우, 60대 중반까지 증가하다가 그 이후 감소하는 경향을 보인다. 기존연구에서 BMI는 60세 때까지 증가하는 경향을 보이며 그 이후 감소한다고 알려져 있으며 이는 본 연구 결과와 유사하였다.<sup>25</sup> 연령증가에 따른 허리둘레는 남자의 경우 서서히 증가하나 일정한 상태를 유지하며, 여성의 경우 70대까지 증가하는 경향을 보인다. 특히 남성에서는 허리둘레가 연령에 따라 지속적으로 증가하는 표본코호트 자료와 비교하여 맞춤형 자료에서는 연령에 따른 증가가 미약하였다. 이 같은 차이는 표본코호트와 맞춤형 자료와의 차이에 기인할 것으로 사료된다. 즉 표본코호트에서 2-30대 선정에 있어 선택편견과 일정 기간에 걸쳐 반복 측정된 자료를 단면 분석함에 따라 연령과의 관련성이 상대적으로 크게 관찰된 결과라 할 수 있을 것이다.

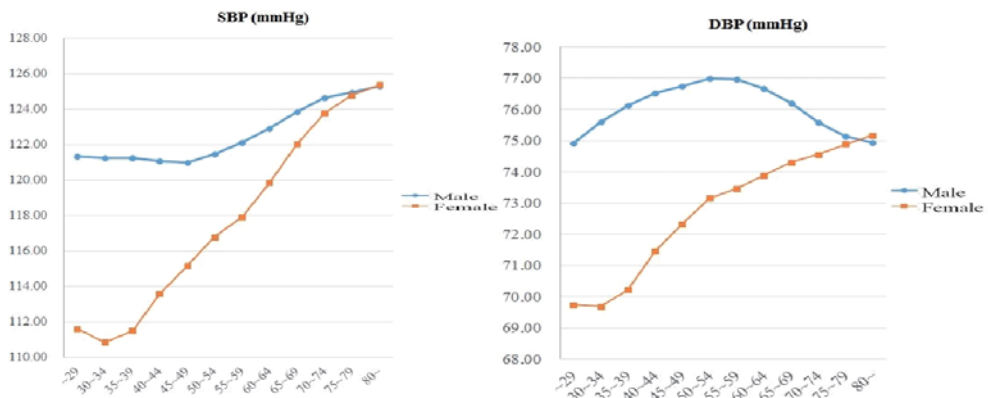


[그림 4-7] 연령에 따른 체중, BMI,허리둘레 변화

## 3) Blood Pressure

혈압은 심혈관 기능을 나타내는 가장 중요한 지표로서 특히 수축기혈압은

연령과 함께 증가하며 심혈관질환의 예측 요소이기도 하다.<sup>26</sup> [그림 4-8]과 같이 수축기혈압은 여성과 비교하여 남자에서는 상대적으로 서서히 증가하는 양상을 보인다. 여성의 경우 30대 중반이후부터 뚜렷하게 증가하는 경향을 보이며, DBP의 경우 여성에서 30대 이후 서서히 증가하는 경향을 보인다. 표본코호트 자료와 비교하여 맞춤형 자료에 있어서는 연령과 혈압과의 관련성이 다소 떨어지는 차이를 보였다. 이 같은 결과는 표본코호트 자료가 일정 기간에 걸쳐 반복 측정된 자료임에도 불구하고 건강노화나이 모델개발을 위해 단면 분석을 시행함에 따라 연령과의 관련성이 상대적으로 크게 관찰된 결과라 할 수 있을 것이다.

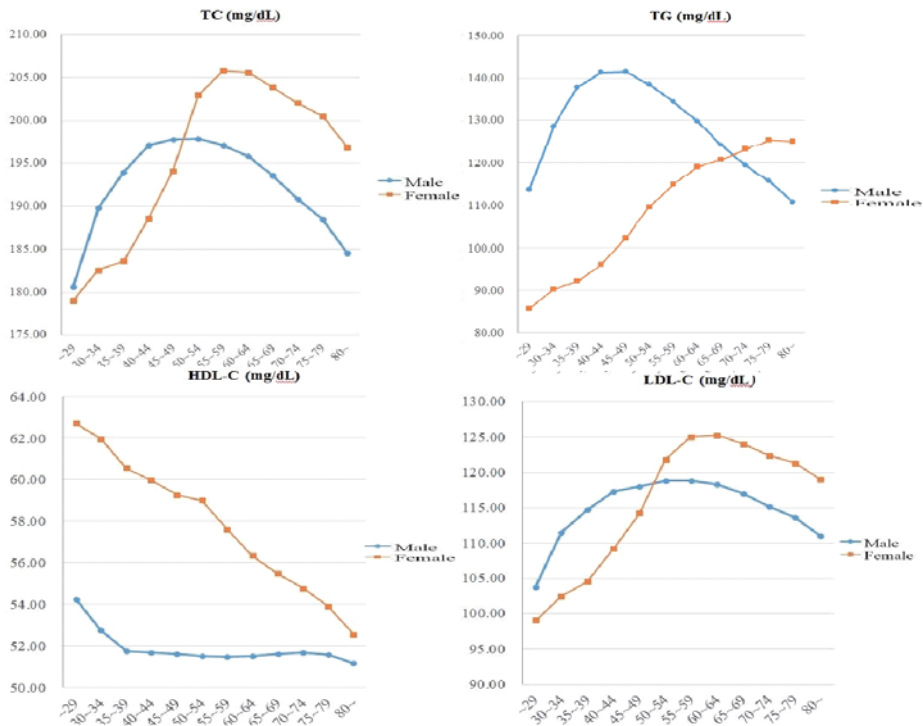


[그림 4-8] 연령에 따른 혈압 변화

#### 4) Lipid profile

연령증가에 따른 총콜레스테롤 변화는 [그림4-9]과 같이 60세 전까지 여성이 남성보다 낮은 값을 나타내나 60세 이후 급격하게 증가하는 경향을 보이며 그 이후로 서서히 감소한다. 남녀 모두 저밀도콜레스테롤은 증가 경향을 보인다. 특히 여성에서 급격한 증가 양상을 보여 55세 이후에는 남성보다 더 높은 값을 나타내며 60대 이후 남녀 모두 감소하는 양상을 나타낸다. 중성지방은 여성에서 연령증가에 따라 급격하게 증가하여 70대 이후에는 남성보다 높은 값을 나타낸다. 고밀도콜레스테롤은 여성에서는 급격한 감소를 보였으나 남성에서는 30대 중반이후 큰 변화가 없었다. 연령에 따른 혈중지질 변화는 기존 Framingham 연구에서 다음과 같이 보고하였다. 총콜레스테롤은

50대 이후 같은 연령대의 남성보다 여성에서 더 높은 수치를 보인다고 한다.<sup>27</sup> 이 같은 결과는 본 연구결과와 유사하였다. 중성지방은 여성보다 남성에서 빠르게 증가하는 것으로 알려져 있으며(Ericsson et al., 1991),<sup>28</sup> 남성에서는 점차 증가하여 40-50세 사이 최고치에 도달한 후 약간 감소한다고 한다. 그러나 여성에서는 전 생애에 걸쳐 증가한다고 알려져 있으며 이 같은 결과는 본 연구 결과와 유사하였다. 고밀도콜레스테롤은 여성에서는 평생 동안 안정적 수치를 유지하다 폐경 이후에는 낮아지는 경향이 있음을 보고하고 있다(Kreisberg and Kasim, 1987).<sup>29</sup> 20세 이후 연령에 따라 남녀 모두 저밀도콜레스테롤의 점진적인 증가를 보고하였다(Abbott et al., 1983).<sup>30</sup> 또한 저밀도콜레스테롤은 여성에서 폐경기 이후 호르몬 변화에 따라 급격한 증가하는 양상을 보인다고 한다.<sup>31</sup>

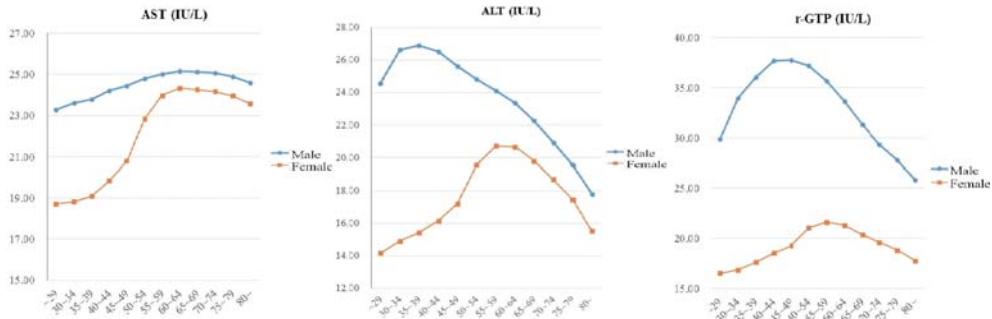


[그림 4-9] 연령에 따른 lipid profile 변화

5) Liver function test

연령증가에 따른 AST의 변화는 [그림4-10]과 같이 남성에서는 서서히 증가하게 되며 여성의 경우 40-60대 급격하게 증가하는 경향을 보인다. ALT 경

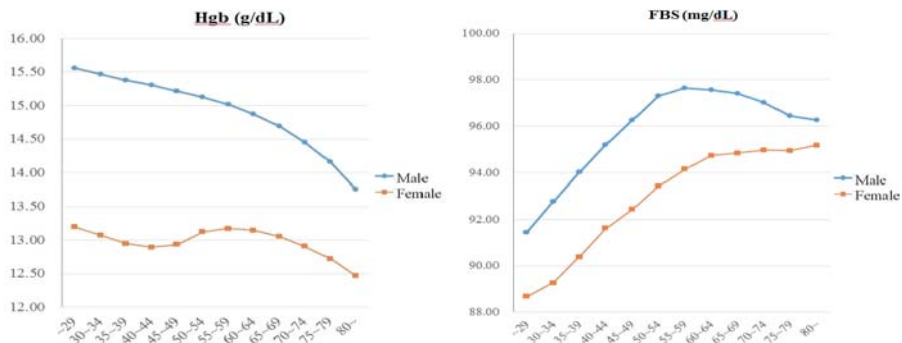
우, 연령증가와 함께 감소하며 여성은 60대까지 증가하다가 감소하는 경향을 보인다.  $\gamma$ -GTP는 남성이 여성보다 높은 값을 보이며 남녀 모두에서 서서히 증가하다 서서히 감소한다. 기존 연구에 따르면  $\gamma$ -GTP는 연령이 증가함에 따라 증가하며, 혈청 aminotransferase는 정상 수준을 지속적으로 유지하는 것으로 보고된 바 있다.<sup>32</sup> 이 같은 결과는 본 연구결과와 유사하였다.



[그림 4-10] 연령에 따른 간기능 변화

6) Hgb, fasting glucose

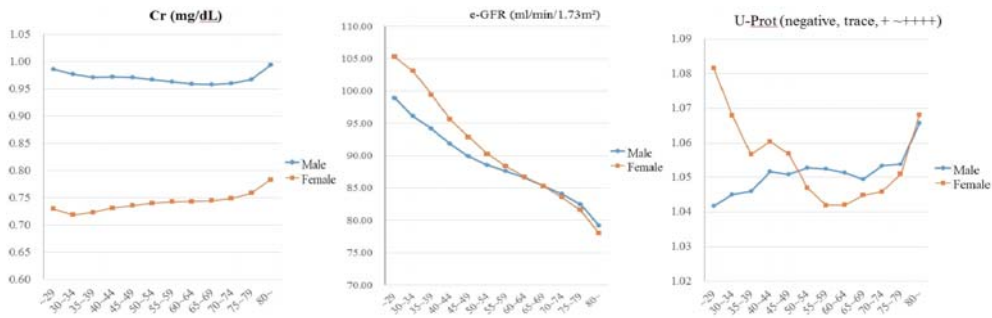
[그림4-11]과 같이 혈색소는 연령증가와 함께 남녀 감소하는 경향을 나타내며 여성보다 남성에서 감소 변화가 더 크고 유의하였다. 공복혈당은 남녀 모두에서 증가하는 경향을 보이며 남성보다 여성에서 보다 급격하게 증가하였다.



[그림 4-11] 연령에 따른 Hgb, FBS 변화

7) renal function test

남녀 모두에서 혈중 크레아티닌은 미세한 증가 추세를 보였으나 유의한 변화는 없었다. 그러나 사구체 여과율은 연령증가에 따라 급격하게 감소하였다.

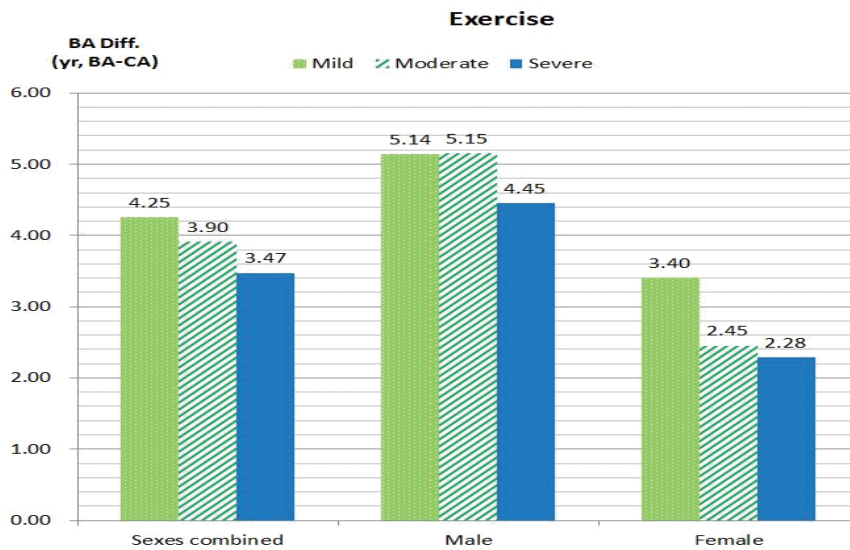


[그림 4-12] 연령에 따른 신기능 변화

## 2. 건강행태별에 따른 건강노화나이

### 1) 활동량에 따른 건강노화나이

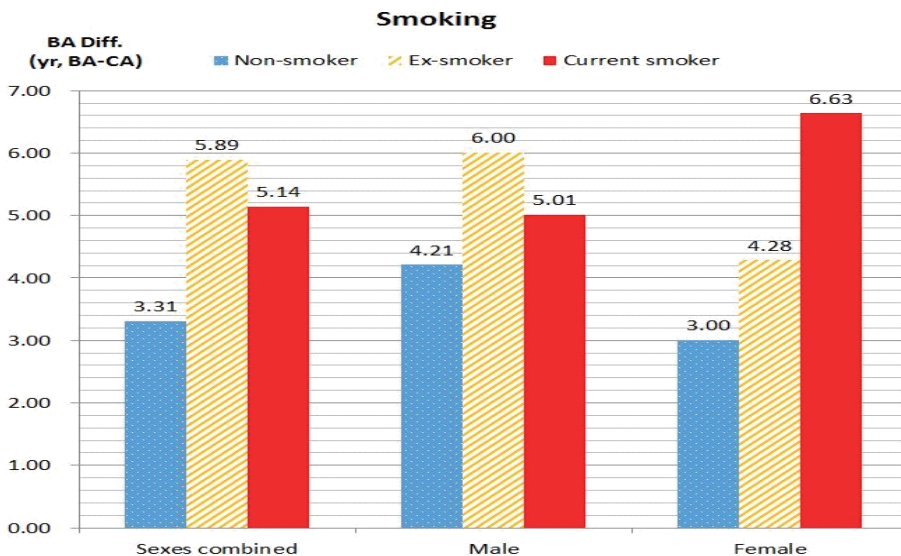
[그림4-13]과 같이 남녀 모두 활동량이 많을수록 건강노화나이는 더 좋았으며 건강노화나이와 주민등록나이와의 차이가 적었다. 특히 경도 활동군과 고도 활동군의 건강노화나이와 주민등록나이와의 차이는 여자에서 상대적으로 크게 나타났다. 기존 건강 체력측정에 따른 건강노화나이와 실제나이의 차이를 규칙적으로 운동하지 않으나 건강한 군, 지속적으로 운동하는 군, 환자군 간의 비교 연구에서 지속적으로 운동하는 군의 건강노화나이와 실제나이의 차이가 14살 정도 유의하게 적게 나타났다.<sup>33</sup>



[그림 4-13] 활동량에 따른 건강노화나이 차이

## 2) 흡연에 따른 건강노화나이

[그림4-14]과 같이 남녀 모두 비흡연군에 비해 흡연군에서 건강노화나이와 주민등록상 나이 차이가 크게 증가하는 것으로 나타났다. 여성에서는 비흡연군과 비교하여 과거흡연군과 현재흡연군 순으로 차이가 뚜렷히 나타났다. 그러나 남성의 경우 과거흡연군에서 건강노화나이 차이가 가장 크게 나타났다. 이 같은 결과는 남성 현재흡연군 가운데 상대적으로 건강 상태가 좋지 않거나 노화 정도가 심각한 대상자들이 더 많이 포함되어 국민건강보험공단 건강검진을 제대로 시행하지 못했을 가능성과 설문지 평가에 따른 응답의 부정확성 등을 고려해 볼 수 있을 것이다. 건강노화나이측정과 사망 예측에 관한 기존연구에서는 흡연에 따른 사망위험도 비교 분석결과 비흡연군에 비해 현재흡연군, 과거흡연군 순으로 사망위험도가 더 높은 것으로 보고하였다<sup>34</sup>.



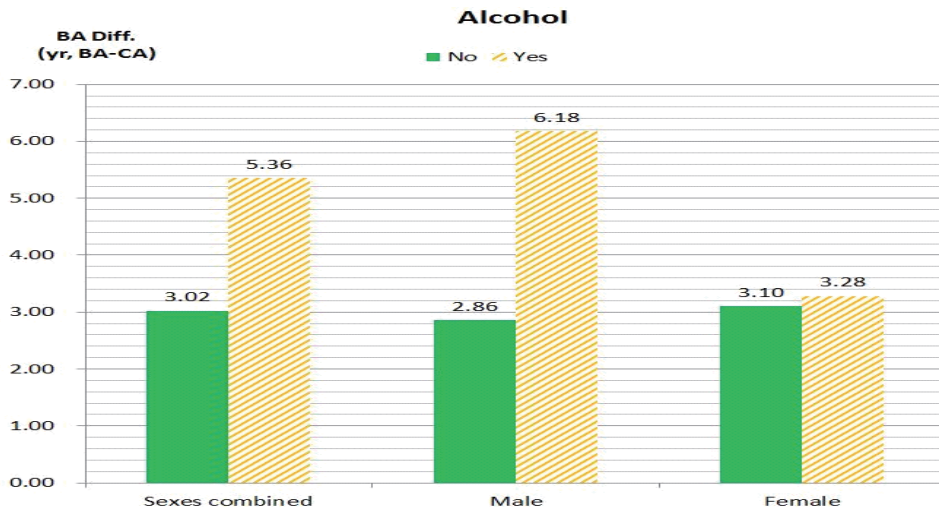
[그림 4-14] 흡연에 따른 건강노화나이 차이

## 3) 음주에 따른 건강노화나이

[그림4-15]과 같이 남녀 모두 음주군에서 건강노화나이와 주민등록상 나이 차이가 큰 것으로 분석되었다. 음주유무에 따른 나이 변화는 여성보다 남성에서는 더 큰 차이를 나타냈다. 음주습관은 주요 생활습관 질환 가운데 고혈압과 당뇨병의 중요한 위험인자로 알려져 있다. 기존 고혈압 혹은 당뇨병 환



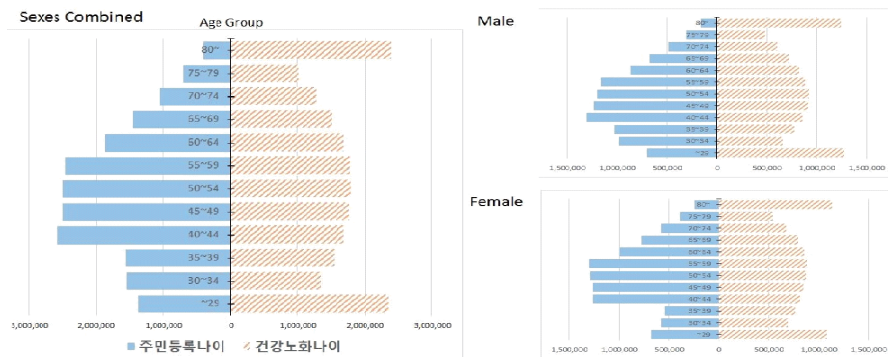
자군의 건강노화나이 연구에 따르면 만성질환군의 건강노화나이가 실제나이보다 높았다고 한다. 고혈군의 건강노화나이가 실제나이보다 4.9세 높았고 당뇨병에서는 7.3세 높았다고 보고하였다.<sup>35</sup> 이 같은 결과는 고혈압과 당뇨병의 주요 위험인자로 알려진 음주가 건강노화나이에도 영향을 미칠 수 있음을 추정할 수 있는 근거가 될 것이다.



[그림 4-15] 음주에 따른 건강노화나이 차이

### 3. 건강노화나이 등급화

#### 1) 연령대 별 분포: 주민등록나이 vs 건강노화나이

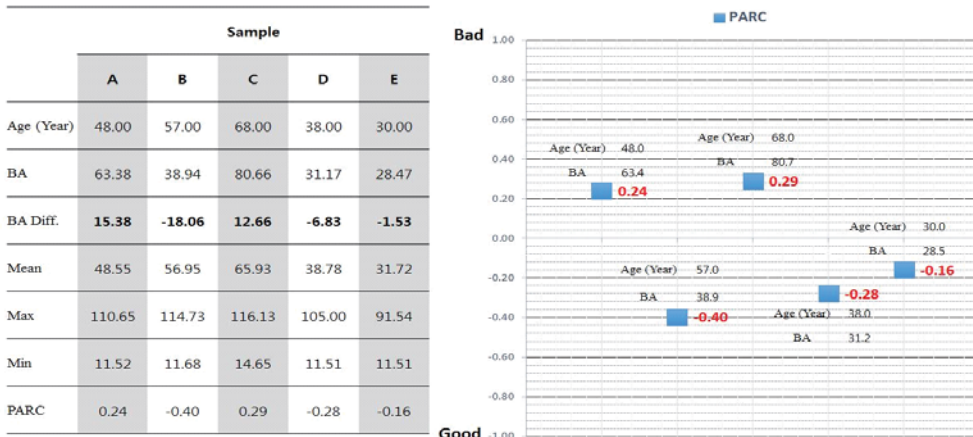


[그림 4-16] 연령대 별 분포: 주민등록나이 vs 건강노화나이

주민등록나이와 건강노화나이별로 연령대별 분포를 분석하였다. 건강노화나이의 20대와 80대에서 분포가 주민등록나이 분포보다 확연히 많음을 확인할 수 있었다. 이 같은 결과는 건강노화나이가 매우 좋거나 매우 나쁜 대상자들이 건강노화나이 양 끝에 모였기 때문으로 설명할 수 있을 것이다. 또한 일정부분 건강노화나이의 모델의 체계적 오차에 따른 결과에 기인할 것으로 생각된다.

2) 건강노화나이에 따른 등급화 예제

[그림4-17] 과 같이 건강노화나이와 주민등록나이의 차이를 PARC index를 사용하여 제시하였다. PARC index는 -1(좋은)~ 1(나쁜) 범위 내로 상대적 비교와 등급을 제시하며 -1에 가까울수록 건강노화나이와 주민등록상 나이 차이가 적음으로 지정하였다.



[그림 4-17] 건강노화나이에 따른 등급화 예제

# 제5장

## 결론

제 1 절 결론	83
제 2 절 제언	88

# 제5장 결론

## 제1절 결론

### 1. 암 질환 예측

#### 1) 위암 예측 연구

본 연구에서는 건강보험공단 건강검진자료를 이용한 인구기반 위암 위험 예측 모델을 개발을 목적으로 과거에 건강검진자료를 기반으로 개발된 위암 위험 모형의 활용성을 점검하였다. 기존 모형은 10년 전 자료를 근거로 생성된 모델이므로 현재의 인구자료에 활용하는데 한계가 있다. 따라서 본 연구에서는 선행 위암 위험 발생 예측 모델 연구와 동일한 접근 방법으로 최신 자료를 이용하여 10년 위암 발생 예측 모델을 구현 하였다. 향후 건강검진 수검자들의 검진결과를 기반으로 위험인자를 측정하여 위암 위험 발생 예측 모델을 구축하고 이용자로 하여금 건강IN 서비스를 통해 위암 발생 위험요소를 파악하고 건강한 생활습관의 변화를 유도할 수 있도록 하는데 그 의의가 있다고 할 수 있다. 본 연구에서는 결과 선행 연구와 유사하게 고령, 저체중 혹은 과체중, 가족력이 있는 경우, 음주량이 높은 경우, 흡연을 과거나 현재에 하는 경우에서 위암의 위험이 증가하는 것으로 나타났다. 그러나 기존과 동일한 모형을 이용한 10년 위암 발생 위험을 예측하기에는 한계가 있는 것을 알 수 있었다.

#### 2) 대장암 위험 예측 연구

대장암 발생위험모형을 만들기 위해 우선적으로 대장암 발생위험요인을 찾았다. 선택된 요인들을 이용하여 최종 대장암 발생위험모형을 구축할 계획이었으나, 클라우드 시스템을 이용하여 분석을 진행하다 보니, 내·외부 검증 을 위해 분석하는 SAS 매크로 구동의 핵심 라이브러리 설치가 되지 않은 상

태여서 발생위험모형을 구축하지는 못하였다. 상기 문제는 추후 시스템의 지원을 받아 진행하도록 할 계획이다.

### 3) 폐암 위험 예측 연구

폐암 모델 위험 예측 개발을 통한 주요 위험 변수를 추후 시범사용에 적용할 예정이다. 추후 보다 나은 모델 개발을 위하여, 함수 구축을 위한 J코드 분류를 위해 검진 전 5년 이내 연도별 진료코드를 받은 6,811,894명 중 진료 코드별 N수와 임상적 의미를 고려하여 J43, J44, J60, J61, J62, J63, J65, J84에 대하여 아래와 같이 3 group으로 범주화하였으며, 다음과 같이 category를 구분하여 분석해 볼 예정이다.

- ① J43 / J44 / J60-70 / J84
- ② J43+J44 / J60-70 / J84

<표 5-1> 폐암 예측 질병 카테고리

코드	구분
J43 폐기종(Emphysema) J44 기타 만성 폐쇄성 폐질환(Other chronic obstructive pulmonary diseases)	만성 하부호흡기질환(J40-J47) Chronic lower respiratory diseases
J60 탄광부진폐증(Coalworker's pneumoconiosis) J61 석면 및 기타 광섬유에 의한 진폐증(Pneumoconiosis due to asbestos and other mineral fibers) J62 실리카를 함유한 먼지에 의한 진폐증(Pneumoconiosis due to dust containing silica) J63 기타 무기물먼지에 의한 진폐증(Pneumoconiosis due to other inorganic dusts) J65 결핵과 연관된 진폐증(Pneumoconiosis associated with tuberculosis)	외부요인에 의한 폐질환(J60-J70) Lung diseases due to external agents
J84 기타 간질성 폐질환(Other interstitial pulmonary diseases)	주로 간질에 영향을 주는 기타 호흡기질환(J80-J84) Other respiratory diseases principally affecting the interstitium

또한 진료 DB를 이용하여 검진 전 5년 이내 주상병 및 부상병이 폐암 관련 호흡기 질환(J60, J61, J62, J63, J64, J65, J81, J84)으로 입원 1회 이상 또는 외래

3회 이상 진료를 받은 사람들을 구분한 결과 폐질환 관련 DB는 아래와 같이 진단명 기준으로 예측 모델 개발식에 적용할 예정이다.

생존분석(cox proportional hazard)의 경우 폐암 검진 주기(폐암 검진 권고 주기 1년) 및 임상적 의미를 고려했을 때 연간 hazard ratio가 reasonable 하며, 추가적으로 5년 단위도 고려하기로 하였다.

## 2. 당뇨 질환 예측에 따른 처방메세지 개발

기존에 개발한 당뇨 위험 예측 모델을 통하여 대국민용 당뇨 위험 예측 프로그램 알고리즘을 개발하였다. 당뇨 위험 예측식을 추정하고 이를 이용하여, 이용자들의 개인의 각 당뇨 위험 요인의 값을 통하여 개인별 10년 당뇨 위험 예측도를 추정하고 제시하였다. 이용자들의 개인의 당뇨 위험도의 교정 가능한 요인을 제시받고, 각 위험 요인별 생활습관 교정 메시지를 제공받을 수 있도록 하였다.

## 3. 심뇌혈관 질환 예측

본 연구에서는 국민 건강 보험 공단의 건강 검진 자료를 이용하여 비혈액검사적 방법을 통한 심뇌혈관 통합 예측 모델을 개발하였다. 금번 개발된 모델의 변별력은 남성이 검증군에서 0.752, 여성이 0.788로 기존 발표된 해외 연구에 비하여 여성의 경우 다소간 예측력이 떨어진다고 할 수 있겠으나 좋은 변별력을 보여주었다. 또한 지난 건강인 연구에서 혈액검사 결과를 사용하였던 모델에 비하여도 별다른 차이가 나지 않는 성능이다.<sup>36</sup>

<표 5-2> 기존에 발표된 심뇌혈관 예측 모델의 변별력 (C-index)

	Male	Female
혈액검사적 방법을 이용한 모델		
Pooled cohort equation	0.744	0.801
Korean heart study	0.764	0.815
비혈액검사적 방법을 이용한 모델		
NHANES II cohort model	0.783	0.831

비혈액검사적 방법을 이용한 심뇌혈관 예측 모델은 본 연구가 처음이 아니며 기존의 비혈액검사적 방법을 이용한 심뇌혈관 예측 모델은 혈액 검사 등의 비용 부담이 어려운 개발도상국에서 저비용으로 심뇌혈관 질환을 예측하는데 사용되도록 개발된 바 있다.<sup>37</sup> 하지만, 이미 개발도상국을 벗어난 국내 실정에서는 단순히 비용 절감을

넘어 의료기관에 방문하여 검사를 받거나 건강검진을 받지 않아도 건강인 서비스 이용자들이 손쉽게 본인의 심뇌혈관 발생 위험도를 계산할 수 있도록 하여 서비스 이용의 진입 장벽을 낮추는 효과를 나타낼 것으로 기대한다. 또한 건강인 서비스 이용자들이 혈액검사 없이 반복적으로 본인의 질병 위험도의 변화를 계산할 수 있게 하여 생활습관 변화에 즉각적인 피드백을 함으로서 보다 적극적인 생활습관 변화를 유도할 수 있는 효과를 나타낼 수 있는 부가적인 효과도 기대한다. 이는 우리나라의 높은 인터넷 및 스마트폰 보급률을 고려하였을 때 기존의 모델과 비교하여 스마트 건강관리에 좀더 적합한 모델이 될 수 있을 것으로 생각된다.

#### 4. 건강노화나이

##### 1) 노화에 따른 생체지표 변화

남자에서 연령 증가에 따라 증가하는 생체지표로서는 허리둘레, 수축기혈압, 이완기혈압, 공복혈당, 총콜레스테롤, 저밀도콜레스테롤, 요단백, 지오티가 있었다. 반대로 연령 증가에 따라 감소하는 생체지표로서 신장, 체중, 체질량지수, 혈색소, 중성지방, 고밀도콜레스테롤, 혈중크레아티닌, 사구체여과율, 지피티, 감마지피티가 있었다. 여자에서 연령 증가에 따라 증가하는 생체지표로서는 체질량지수, 허리둘레, 수축기혈압, 이완기혈압, 혈색소, 공복혈당, 총콜레스테롤, 중성지방, 저밀도콜레스테롤, 혈중크레아티닌, 지오티, 지피티, 감마지피티가 있었다. 감소하는 생체지표로서는 신장, 체중, 고밀도콜레스테롤, 요단백, 사구체여과율이 있었다.

##### 2) 건강행태에 따른 건강노화나이

###### ① 활동량에 따른 건강노화나이

남녀 모두 활동량이 많을수록 건강노화나이는 더 좋았으며 건강노화나이와 주민등록나이와의 차이가 적었다. 성별에 따른 건강노화나이와 주민등록나이와의 차이는 정도 활동군(남 5.1세, 여 3.4세)에서는 1.7세, 중등도 활동군(남 5.2세, 여 2.5세)은 2.7세, 고도 활동군에서(남 4.5세, 여 2.3세)로 2.2세로 분석되었다. 특히 정도 활동군과 고도 활동군과의 건강노화나이와 주민등록나이와의 차이는 여자에서 상대적으로 크게 나타났다.

###### ② 흡연에 따른 건강노화나이

남녀 모두 비흡연군에 비해 흡연군에서 건강노화나이가 나쁜 것으로

분석되었다. 여자에서는 현재 흡연군(6.6세), 과거 흡연군(4.3세), 비흡연군(3.0세) 가운데 현재 흡연군의 건강노화나이 차이가 6.6세로 가장 많은 것으로 분석되었다. 그러나 남자에서는 현재 흡연군(5.0세), 과거 흡연군(6.0세), 비흡연군(4.2세) 가운데 과거 흡연군의 건강노화나이 차이가 가장 큰 것으로 분석되었다.

### ③ 음주에 따른 건강노화나이

남녀 모두 음주군에서 건강노화나이가 나쁜 것으로 분석되었다. 특히 남자에서는 음주군(6.2세)과 비음주군(2.9세)과의 차이가 3.3세로 여자 음주군(3.3세)과 비음주군(3.1세) 차이 0.2세 보다 가 큰 것으로 분석되었다.

### 3) 건강노화나이 등급화

PARC Index를 사용하여 건강노화나이와 주민등록나이와의 차이를 -1(좋은)~+1(나쁜) 범위 내에서 비교하고 상대적 등급 예시를 제시하였다. PARC Index에 따른 상대적 비교와 등급 제시를 통해 건강노화나이에 대한 일반 수검자들의 이해를 높이는 또 다른 수단으로 활용하고자 하였다.

### 4) 연구의의

2016년 연구과제에서는 2002년부터 2013년 국민건강보험공단 표본코호트 자료를 사용하여, 건강과 노화를 평가하는 과학적 지표로 활용될 수 있는 객관적 근거를 제시하였다. 건강노화나이가 주민등록나이에 비해 많을수록 사망위험성이 높고 이와 같은 결과는 건강노화나이가 사망위험성을 예측 할 수 있다는 객관적 근거를 제시하였다. 건강노화나이가 주민등록나이에 비해 많을수록 주요 질환발생률이 높고 이와 같은 결과는 건강노화나이가 주요 질환의 발생위험성을 예측 할 수 있다는 객관적 근거를 제시한 바 있다.

2017년 연구과제에서는 다음과 같은 3가지 사항을 제시하였다. 첫째, 개발된 건강노화나이 모델을 2014-15년 국민건강보험공단 전수자료에 적용하여, 국가 적정 표준화를 제시하였다. 건강노화나이를 구성하는 생체지표들의 노화에 따른 증감 변화를 분석하여 성별, 연령대 별 현황을 확인하였으며, 생활습관에 따른 건강노화나이 분석을 통해 건강노화나이의 외적 유용성과 타당성을 검증하였다. 둘째, 건강노화나이 등급화를 통해서 이는 추후 "나이"에 따른 상호 소통과 비교 이외에 또 다른 동기부여 수단이나 관리목표로 활용될 수 있을 것이다. 또한 등급화에 따른 직관적이고 간략한 비교를 통해



건강노화나이의 원활한 적용과 활용에 도움을 줄 수 있을 것이다. 셋째, 건강노화나이의 국가 적정 표준화와 등급화에 따른 고도화 분석을 통해서 실제적으로 활용되고 적용될 수 있는 근거를 제시하였다. 일반 수검자들의 만성 질환 예방과 건강 증진에 실제적 도움이 될 수 있는 지표로서, 추후 이는 건강검진 결과를 통해 개인의 건강 상태와 노화 정도를 종합적으로 제시하고 개별 수검자들의 결과를 서로 쉽게 이해하고 소통하며 올바른 생활습관 관리를 위한 구체적 관리목표 제공할 수 있을 것이다.

## 제2절 제 언

이 연구는 2014년부터 시작된 건강iN의 4차년도 후속 연구로써 이 연구 결과를 이용하여 추후 대국민 맞춤형 건강 서비스를 개발하고자 하였다. 이 연구의 과학적 기반 근거 위에, 각 국민 개개인의 눈높이에 맞춘 적절한 건강관리 서비스의 내용 및 전달방법이 요망된다.

### 1. 위암 예측 연구

#### 1) 연구의 한계 및 향후 연구 방향

본 연구는 남성, 여성 모델에서 모두 높은 판별력을 보인 것과는 달리 calibration에서 개발된 모델의 예측값과 실제값에 차이를 보이는 한계를 가지고 있다. 차이에 대한 원인으로는 본 연구에 포함된 위험 인자들만으로는 위암을 예측하기 어려운 것으로 보이며, 암은 오랜 시간에 걸쳐 발생하는데 비해 관측기간이 길지 않고 나이, 흡연과 가족력을 제외한 다른 위험 인자들에서는 분석 결과 위험비가 1에 가까워 범주 간 차이가 나타나지 않았기 때문에 위암을 예측하는데 한계를 가지고 있는 것으로 볼 수 있다. 염분 섭취 및 육류 섭취 등의 구체적인 식이 관련 변수들과 *Helicobacter pylori*, Peptic ulcer 와 같은 위암 발생에 영향을 미치는 위험 요인들을 추가적으로 고려한 후속 연구가 필요할 것으로 사료된다.

### 2. 당뇨 질환 예측에 따른 처방 메시지 개발

전세계와 마찬가지로 한국에서도 당뇨의 유병률이 지속적으로 증가하고 있고, 당뇨는 주요 심혈관계 사망원인에 기여하고 있으므로, 당뇨 위험을 낮추기 위한 대국민용

교육 프로그램이 필요하다. 이러한 필요성을 배경으로, 본 연구에서 개발한 당뇨 위험 예측 모델 및 알고리즘을 이용하여, 건강iN의 건강위험 예측 프로그램에 신규로 탑재하는 실질적, 행정적 프로세스가 필요하다.

### 3. 심뇌혈관 질환 예측

본 연구는 전술한 바와 같이 10년 질환 발생 위험도를 예측하기 위하여 2002년과 2003년에 건강검진을 받은 인구를 대상으로 수행되었다. 당시에는 콜레스테롤의 하위 구분을 하지 않고 총콜레스테롤 수치만을 측정하였으며, 허리둘레, 혈청 크레아티닌 수치 등 모델의 예측력을 좀 더 증가시킬 수 있는 인자들에 대한 측정이 이루어지지 않았다. 기존 임상진료 지침에 사용되고 있는 해외의 모델의 경우 이러한 변수가 포함되어 있어 본 연구에서 개발한 모델은 기존 모델과의 직접적인 성능 비교가 불가능하다. 향후 2008년 이후 건강검진 대상자들의 자료가 확보된다면 이러한 부분을 고려한 새로운 모델의 개발이 필요할 것으로 생각된다. 추가적으로 위험 예측도 계산을 이용한 생활 습관 교정의 효과는 전향적인 무작위 대조연구에서 입증된 바 있으나 비혈액적 모델을 이용한 반복적인 위험도 측정으로 인한 생활 습관 교정의 동기부여 효과는 아직 전향적인 무작위 대조연구에서 입증된 바가 없다. 따라서 본 연구에서 개발된 모델의 실질적인 효과 측정은 추가적인 전향적인 무작위 대조연구에 의한 입증이 필요함을 밝힌다.

### 4. 건강노화나이

#### 1) 국민건강보험공단 건강검진 측정항목

개인의 건강과 노화를 종합적으로 평가하는 측면에서 건강검진 측정항목이 미리 정해져 있기에 제한적일 수밖에 없다. 전수자료에 적용된 건강노화나이 모델은 표본코호트 자료에 기초하여 다소 엄격한 이론적 근거와 통계분석 기준을 적용하여 개발되었다. 따라서 임상적으로 중요 질환의 평가지표(혹은 생체지표)들이 일부 배제된 부분이 있을 것으로 사료된다. 따라서 국민건강보험공단 건강검진의 측정항목만으로도 건강 상태와 노화 정도를 정확하게 설명할 수 있는 건강노화나이 모델을 제시하였지만 앞으로도 이를 개선시키기 위한 지속적 연구가 뒤따라야 할 것이다.

실제 적용과 활용 시 고려사항으로서 노화와의 연관성은 다소 미약하더라도 생활습관과의 관련성이 잘 알려져 있는 생체지표를 포함한 실제적인 모델

적용이 필요할 것으로 사료된다. 특히, 남자 건강노화나이는 모델에 포함된 생체지표(신장, 허리둘레, 수축기혈압, 공복혈당, 혈색소, 사구체여과율, 지오티)이외에 이완기혈압, 혈중지질, 감마지피티 등과 같은 생체지표가 포함된 모델을 적용하여 개별 수검자들의 건강과 노화 관리에 보다 실제적인 도움을 주는 것이 필요할 것이다. 이를 통해 기존 생활습관 질환의 적극적 관리와 함께 올바른 생활습관을 유지하고 개선시킬 수 있는 구체적 동기부여 수단과 종합 목표지표로서 활용될 수 있을 것으로 사료된다.

## 2) 서비스 도입과 활용

과학적 근거와 검증 측면에서 주요 질환 이외의 여러 질환, 기능저하, 잔여수명 등과 연관된 다양한 연구가 지속적으로 진행되어야 할 것이다. 종합지표, 소통지표, 관리지표로서 건강노화나이의 적용 효과, 실제 활용의 편리성, 수검자 만족도 등과 같은 서비스 적용에 관한 연구를 통해 다양한 근거를 제시하는 노력이 필요할 것이다. 건강노화나이 관련 자료측면에서는 측정된 건강노화나이의 종합분석, 생체지표 변화에 따른 설명, 올바른 생활습관 유지 및 개선에 필요한 구체적 실행방법 등과 같은 다양한 관련 자료 개발을 통해 수검자들이 관리목표를 설정하고 이를 이해하고 실천하는데 도움이 될 수 있어야 할 것이다.

## 3) 등급화의 실제 적용

예시에서 제시된 -1(좋음)~+1(나쁨) 표현방식 이외에 생활 속에서 흔히 접하고 친숙한 5등급(매우 좋음, 좋음, 평균, 나쁨, 매우 나쁨) 혹은 10등급(1등급~10등급) 등의 구체적 표현 방식을 고려해 볼 수 있을 것이다.

참고문헌

## 참고문헌

1. Bang Woll Eom et al. Prediction Model for Gastric Cancer Incidence in Korean Population. PLOS ONE. 10(7): e0132613.
2. Yoo Mi Park et al. Clinical risk assessment for gastric cancer in asymptomatic population after a health check-up: An individualized consideration of the risk factors. Medicine (216) 95:44(e5351).
3. Teng-Yu Lee et al. A Tool to Predict Risk for Gastric Cancer in Patients With Peptic Ulcer Disease on the Basis of a Nationwide Cohort. Clinical Gastroenterology and Hepatology 2015;13:287-293.
4. Zhensheng Wang et al. Composite protective lifestyle factors and risk of developing gastric adenocarcinoma: the Singapore Chinese Health Study. British Journal of Cancer (2017) 116, 679-687.
5. G. Buckland et al. Healthy lifestyle index and risk of gastric adenocarcinoma in the EPIC cohort study. International Journal of Cancer: 137, 598-606 (2015).
6. Shin A, Joo J, Yang HR, et al. Risk prediction model for colorectal cancer: National Health Insurance Corporation study, Korea. PLoS One. 2014 Feb 12;9(2):e88079.
7. Park S, Nam BH, Yang HR, et al. Individualized risk prediction model for lung cancer in Korean men. PLoS One 2013;8(2):e54823.
8. Gray, Eoin P., et al. "Risk Prediction Models for Lung Cancer: A Systematic Review." Clinical lung cancer (2015).
9. World Health Organization. Global report on diabetes. World Health Organization, 2016.
10. Federation, International Diabetes. "IDF diabetes atlas." Brussels: International Diabetes Federation (2015).
11. 대한당뇨병학회, <http://www.diabetes.or.kr/general/index.html>

12. Ha, K. H., & Kim, D. J. (2015). Trends in the diabetes epidemic in Korea. *Endocrinology and Metabolism*, 30(2), 142–146.
13. Song, S. O., Lee, Y. H., Kim, D. W., Song, Y. D., Nam, J. Y., Park, K. H., ... & Lee, B. W. (2016). Trends in diabetes incidence in the last decade based on Korean National Health Insurance claims data. *Endocrinology and Metabolism*, 31(2), 292–299.
14. World Health Organization. Global report on diabetes. World Health Organization, 2016.
15. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98(9):683-90.
16. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98(9):691-8.
17. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*. 2014;63(25):2889-934.
18. David C Goff J, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014;129(6): S49-47.
19. Crimmins E1, Vasunilashorn S, Kim JK, Alley D. et al. Biomarkers related to aging in human populations. *Adv Clin Chem*. 2008;46:161-216.
20. Duggirala R1, et al. Genetic determination of biological age in the Mennonites of the Midwestern United States. *Genet Epidemiol*. 2002 Aug;23(2):97-109
21. T. Baker, III and Richard L. Sprott (1988). "Biomarkers of aging". *EXPERIMENTAL GERONTOLOGY*. 23 (4-5): 223-239.
22. 남수민 et al. 『우리나라 대장암 발병률의 양상과 위험인자에 대한 연구. 2017』

23. Cline MG, Meredith KE, Boyer JT, Burrows B. Decline of height with age in adults in a general population sample: estimating maximum height and distinguishing birth cohort effects from actual loss of stature with aging. *Hum Biol.* 1989;61(3):415-425.
24. Seidell JC, Visscher TL. Body weight and weight change and their health implications for the elderly. *Eur J Clin Nutr.* 2000;54 Suppl 3: S33-S39.
25. Wells JC, Cole TJ, Treleaven P. Age-variability in body shape associated with excess weight: the UK National Sizing Survey. *Obesity (Silver Spring).* 2008;16(2):435-441.
26. Izzo JL Jr, Levy D, Black HR. Clinical Advisory Statement. Importance of systolic blood pressure in older Americans. *Hypertension.* 2000;35(5):1021-1024.
27. Abbott, R.D., Garrison, R.J., Wilson, P.W., et al., 1983. Joint distribution of lipoproteincholesterol classes: The Framingham Study. *Arterioscler. Thromb. Vasc. Biol.* 3,260-272.
28. Ericsson, S., Eriksson, M., Vitols, S., Einarsson, K., Berglund, L., Angelin, B., 1991. Influence of age on the metabolism of plasma low density lipoproteins in healthymales. *J. Clin. Invest.* 87, 591-596.
29. Kreisberg, R.A., Kasim, S., 1987. Cholesterol metabolism and aging. *Am. J. Med.* 82,54-60.
30. Abbott, R.D., Garrison, R.J., Wilson, P.W., et al., 1983. Joint distribution of lipoproteincholesterol classes: The Framingham Study. *Arterioscler. Thromb. Vasc. Biol.* 3,260-272.
31. Liu et al. Aging and dyslipidemia :A review of potential mechanisms. *Ageing Res Rev.* 2015 Jan;19:43-52.
32. Tietz NW, Shuey DF, Wekstein DR. Laboratory values in fit aging individuals: sexagenarians through centenarians. *Clin Chem.* 1992; 38:1167-1185.

33. Lee MS1, Tanaka K et al. The relative utility of health-related fitness tests and skilled motor performance tests 198as measures of biological age in Japanese men. *Appl Human Sci*. 1996 May;15(3):97-104.
34. Bulpitt CJ et al. Mortality according to a prior assessment of biological age. *Curr Aging Sci*. 2009 Dec;2(3):193-9.
35. Nakamura E et al. Evaluation of the progress of aging based on specific biological age as estimated by various physiological functions. *Nihon Eiseigaku Zasshi*. 1982 Feb;36(6):853-62.
36. Jee SH, Jang Y, Oh DJ, et al. A coronary heart disease prediction model: the Korean Heart Study. *BMJ Open* 2014;4(5): 502-15.
37. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. *The Lancet* 2008;371(9616): 923-31.



연구보고서 2017-20-027

## **개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발(4차년도)**

---

<b>발행일</b>	2017년 12월 30일
<b>발행인</b>	강중구
<b>편집인</b>	장호열
<b>발행처</b>	국민건강보험공단 일산병원 연구소
<b>주소</b>	경기도 고양시 일산동구 일산로 100
<b>전화</b>	031) 900-6982~6987
<b>팩스</b>	031) 900-6999
<b>인쇄처</b>	백석기획(031-903-9979)

---



국민건강보험 일산병원 연구소



(우)10444 경기도 고양시 일산동구 일산로 100 (백석1동 1232번지)  
대표전화 031-900-0114 / 팩스 031-900-6999  
[www.nhimc.or.kr](http://www.nhimc.or.kr)

## 2017 NHIS Ilsan Hospital Institute of Health Insurance & Clinical Research

N a t i o n a l H e a l t h I n s u r a n c e S e r v i c e I l s a n H o s p i t a l