

연구보고서 2020-20-018

www.nhimc.or.kr

# 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발(7차년도)

조경희 · 박주현 · 최준호 · 이상현 · 전성연 · 이재광 · 임현선  
김동욱 · 배세진 · 조현순 · 김보현 · 김 번 · 최일주 · 배철영  
서강춘 · 김미영 · 임유리 · 김소현 · 김유미

# NHIS

2020 NHIS Ilsan Hospital  
National Health Insurance Service Ilsan Hospital

국민건강보험

National Health  
Insurance Service

일산병원

Ilsan Hospital

연구보고서

2020-20-018

# 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발(7차년도)

조경희·박주현·최준호·이상현·전성연·이재광·임현선·김동욱  
배세진·조현순·김보현·김번·최일주  
배철영·서강춘·김미영·임유리·김소현·김유미

국민건강보험

National Health  
Insurance Service

일산병원

Isan Hospital

국민건강보험 일산병원 연구소

[저 자]

책임 연구자:	국민건강보험 일산병원 가정의학과	조경희
공동 연구원:	국민건강보험 일산병원 가정의학과	박주현
	국민건강보험 일산병원 가정의학과	최준호
	국민건강보험 일산병원 가정의학과	이상현
	국민건강보험 일산병원 연구분석팀	전성연
	국민건강보험 일산병원 연구분석팀	이재광
	국민건강보험 일산병원 연구분석팀	임현선
	국민건강보험공단 빅데이터실	김동욱
	국민건강보험공단 빅데이터실	배세진
	국립암센터	조현순
	국립암센터	김보현
	국립암센터	김번
	국립암센터	최일주
	메디에이지	배철영
	메디에이지	서강춘
메디에이지	김미영	
메디에이지	임유리	
국민건강보험 일산병원 가정의학과	김소현	
보조 연구원:	국민건강보험 일산병원 교수지원팀	김유미

연구관리번호	IRB 번호
NHIS-2020-1-281	NHIMC 2020-02-002

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

## 머리말

건강IN은 인터넷상 검증되지 않고 무분별하게 서비스 되고 있는 건강정보로부터 국민건강을 보호하기 위하여 인터넷상 검증되지 않고 무분별하게 서비스 되고 있는 건강정보로부터 국민건강을 보호하기 위한 건강정보 전문사이트이다.

본 연구는 7년에 걸쳐서 검증한 신뢰성 있는 건강정보를 제공하도록 건강인 사이트 관련 연구를 지원해 오고 있다. 다빈도 질병/처치/수술/검사/흡연/영양 등 다양한 건강정보를 개발해 왔으며, 공단이 보유하고 있는 국가건강검진자료를 이용한 개인별 맞춤형 건강정보 연구, 비만·흡연자를 위한 맞춤형 생활습관 개선 프로그램 연구, 개인의 위험도 평가 후 건강개선자료를 제공하는 건강나이 알아보기 연구, 대사성질환자를 위한 맞춤형 정보 연구 등을 수행해 왔다.

최근 수년간은 각종 암에 관한 질병 예측 연구와 함께 4차 산업 혁명시대의 맞춤형 정보 관련 연구를 수행하였다. 2019년에는 대사증후군의 표현형과 유전형의 상관성 분석을 통하여 유전체 관련 연구를 수행하였고, 2020년에는 대사증후군 환자 생활 습관 개선을 위하여 IoT를 이용한 Digital Therapeutics(DTx)를 적용하여 미래 의료의 방향을 살펴보았다. 또한 5차례의 건강인 포럼을 통하여, 미래 건강인이 나아갈 방향에 대한 논의를 하였다. PHR, 진료정보교류사업, 의료데이터 이용, 국민건강교육자료 등의 주제와 함께 국가에서 주도하고 있는 4차산업혁명위원회 디지털헬스특위에서의 방향에 관한 토의를 하였다. 개인 주도형 의료데이터 실행 전략은 어디 한군데에 정보를 모으지 않고 개인 의료데이터 보호·보안 및 개인 주도적 접근성·활용이 보장된 My Healthway라는 gate way를 구축하여 의료기관, 보험, 타 기관 정보 등 여러 가지 종류의 정보가 개인사용 동의를 하면 활용이 될 수 있도록 준비되고 가고 있으며, 정보가 필요할 때 언제든지 활용할 수 있도록 진행되고 있다.

국가 PHR 사업에 맞추어 건강인도 신뢰 있는 건강 플랫폼을 위한 지속적인 고도화와 함께 앱을 통한 맞춤형 건강 정보 제공을 위한 지속적인 노력이 경주 된다.

2020년 10월

국민건강보험 일산병원장

일산병원 연구소장

김성우  
오현철

# 목차

요약 .....	1
<b>제1장 서론</b> .....	<b>9</b>
제1절 연구 배경 및 필요성 .....	11
제2절 연구의 목적 .....	12
<b>제2장 이론적 고찰</b> .....	<b>15</b>
제1절 간암 질환 예측 .....	17
제2절 대장암 질환 예측 .....	23
제3절 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과 · 25	
제4절 건강인에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고 · 27	
<b>제3장 연구 내용 및 방법</b> .....	<b>29</b>
제1절 연구내용 .....	31
제2절 세부 과제별 연구 방법 .....	33
<b>제4장 분석결과</b> .....	<b>43</b>
제1절 간암 발생 예측 연구 .....	45
제2절 대장암 발생 예측 연구 .....	67
제3절 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과 연구 .....	88
제4절 건강인에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고 · 93	
<b>제5장 결론 및 제언</b> .....	<b>95</b>
제1절 결론 .....	97
제2절 제언 .....	99
<b>참고문헌</b> .....	<b>103</b>
<b>부록 : 첨부&gt; 포럼 회의록</b> .....	<b>109</b>

## 표목차

<표 2-1> 성별 주요 암 발생 현황, 2017년도 국가암등록통계자료 .....	17
<표 2-2>연령 군별 주요 암발생률: 남녀전체, 2017, 국가암등록통계자료 .....	18
<표 2-3> multi variable cox model for HCC occurrence and CAMD score ...	19
<표 2-4> Factors associated with HCC by cox proportional hazards regression ·	19
<표 2-5> Cox models evaluating risk factors for the ADREESS-HCC model ·	20
<표 2-6> significant risk factor for HCC .....	21
<표 2-7> 기존 간암예측모형에서 활용된 위험요인들 .....	22
<표 2-8> 대장암 예측모형별 위험요인 비교분석 .....	24
<표 2-9> C statistic and Hosmer-Lemeshow type chi-square test for colorectal cancer risk prediction model for development set(8-year risk) and validation set(5-year risk) .....	25
<표 3-1> 맞춤형 DB 세부특성 .....	31
<표 3-2> modified ATP III criteria .....	41
<표 4-1> 일반검진 대상자의 간암 발생에 따른 기본적 특성 .....	46
<표 4-2> 일반검진 대상자의 간암 발생 예측모델 (생존분석) .....	47
<표 4-3> 일반검진 대상자의 간암 발생 예측모델 (로지스틱 회귀분석) .....	48
<표 4-4> 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측모델 (생존분석) .....	53
<표 4-5> 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측모델 (로지스틱 회귀분석) .....	54
<표 4-6> 간암검진 대상자의 간암 발생에 따른 기본적 특성 .....	59

<표 4-7> 간암검진 대상자의 간암 발생 예측모델 (생존분석) .....	61
<표 4-8> 간암검진 대상자의 간암 발생 예측모델 (로지스틱 회귀분석) .....	62
<표 4-9> 일반검진 대상자의 대장암 발생에 따른 기본적 특성 .....	68
<표 4-10> 일반검진 대상자의 대장암 발생 예측모델 (생존분석) .....	69
<표 4-11> 일반검진 대상자의 대장암 발생 예측모델 (로지스틱 회귀분석) .....	70
<표 4-12> 대장암 검진 대상자의 대장암 발생에 따른 기본적 특성 .....	75
<표 4-13> 대장암 검진 대상자의 대장암 발생 예측모델 (생존분석) .....	76
<표 4-14> 대장암 검진 대상자의 대장암 발생 예측모델 (로지스틱 회귀분석) .....	77
<표 4-15> 분변잠혈검사 양성 대상자의 대장암 발생에 따른 기본적 특성 .....	81
<표 4-16> 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 (생존분석) .....	82
<표 4-17> 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 (로지스틱 회귀분석) ..	83
<표 4-18> 내시경 검사 결과 .....	87
<표 4-19> 대장이종조영검사 결과 .....	87
<표 4-20> 연구대상자 특성 .....	88
<표 4-21> 대사증후군 타입 분포 비교 .....	89
<표 4-22> 대사나이 평균 비교 .....	89
<표 4-23> 대사나이 타입 분포 비교 .....	89
<표 4-24> 운동 횟수 평균 비교 .....	90
<표 4-25> 흡연자 분포 비교 .....	91
<표 4-26> 음주자 분포 비교 .....	91
<표 4-27> 음주 횟수 평균 비교 .....	91
<표 4-28> 음주량(잔) 평균 비교 .....	91
<표 4-29> 수면시간 분포 비교 .....	91
<표 4-30> 스트레스 점수 평균 비교 .....	92
<표 4-31> 영양 습관 분포 비교 .....	92
<표 4-32> 걸음 수 평균 비교 .....	92

## 그림목차

[그림 2-1] 국가 암검진 지원사업 .....	23
[그림 3-1] 일반검진자 대상 간암 질환 예측 연구 대상자 선정 .....	34
[그림 3-2] 간암검진자 대상 간암 질환 예측 연구 대상자 선정 .....	34
[그림 3-3] 일반검진자 대상 대장암 질환 예측 연구 대상자 선정 .....	37
[그림 3-4] 대장암검진자 대상 대장암 질환 예측 연구 대상자 선정 .....	38
[그림 4-1] 일반검진 대상자의 생존분석 nomogram .....	51
[그림 4-2] 일반검진 대상자의 calibration plot .....	52
[그림 4-3] 일반검진 대상자의 ROC curve .....	52
[그림 4-4] 일반검진 대상자의 간암 발생 flow chart .....	53
[그림 4-5] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 생존분석 nomogram .....	56
[그림 4-6] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 calibration plot .....	57
[그림 4-7] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 ROC curve ...	57
[그림 4-8] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 flow chart ...	58
[그림 4-9] 간암검진 대상자의 생존분석 nomogram .....	65
[그림 4-10] 간암검진 대상자의 calibration plot .....	66
[그림 4-11] 간암검진 대상자의 ROC curve .....	66
[그림 4-12] 간암 검진 대상자의 간암 발생 flow chart .....	67
[그림 4-13] 일반검진 대상자의 대장암 발생 예측모델 Nomogram .....	72
[그림 4-14] 일반검진 대상자의 대장암 발생 예측모델 Calibration plot .....	73



[그림 4-15] 일반검진 대상자의 대장암 발생 예측모델 ROC curve .....	73
[그림 4-16] 일반검진 대상자의 대장암 발생 flow chart .....	74
[그림 4-17] 대장암 검진 대상자의 대장암 발생 예측모델 Nomogram .....	78
[그림 4-18] 대장암 검진 대상자의 대장암 발생 예측모델 Calibration plot .....	79
[그림 4-19] 대장암 검진 대상자의 대장암 발생 예측모델 ROC curve .....	79
[그림 4-20] 대장암 검진 대상자의 대장암 발생 flow chart .....	80
[그림 4-21] 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 Nomogram .....	85
[그림 4-22] 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 Calibration plot ..	86
[그림 4-23] 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 ROC curve .....	86

요약





## 요약

### 1. 연구 배경 및 목적

본 연구는 2014년도부터 시작된 『개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에 이은 후속 7차년도 연구이다. 『2020년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 간암, 대장암의 일반대상자 및 암 검진 대상자 각각에 대하여 위험요인을 분석해보고 예측모형을 개발하였다. 국가검진대상자 중에서 사후관리 프로그램 참여 동의를 받은 맞춤형 모바일 건강관리 프로그램의 효과 검증과 함께 생활습관과 건강상태 개선 효과를 확인하고 이를 통해 건강생활 습관을 유도하여 질병 발생의 예방과 삶의 질 향상에 도움이 되고자 하였다. 건강in-전문가포럼에서는 의무기록에서의 EMR & EHR 데이터, PHR의 현황과 방향, 진료정보교류 현황, 국민건강교육자료 등의 주제에 대하여 현황과 방향을 살펴보고 앞으로의 정보화 사회에서의 건강인의 방향을 논의하였다.

### 2. 연구 내용 및 방법

간암 질환 예측 연구 및 대장암 질환 예측 연구에서는 건강보험공단 빅데이터의 맞춤형 자료를 이용하여 진행하였다. 간암 질환 예측 연구는 2009년 국가일반검진 대상자와 간암검진 대상자 전수를 대상으로 하여 2004년에서 2008년간 간암 진단코드를 받은 대상자는 제외하고, 2018년까지 총 10년간의 간암 발생여부를 조사하였다. 그리고 간암 검진자의 위험군으로 선정한 HBV, HCV, 기타간염, 간경화를 제거한 인원수를 대상으로 데이터를 구축하였다. 대장암질환 예측 연구에서는 2009년 국가일반검진 대상자와 대장암검진 대상자 전수를 대상으로 2004년에서 2008년간 대장암 진단코드를 받은 대상자는 제외하고, 2018년까지 총 10년간의 간암 발생여부를 조사하였다. 그리고 대장암검진 대상자 중 분변잠혈검사 양성자 중에서 2018년까지 총 10년간의 대장암 발생여부를 조사하였다.

사후관리 프로그램을 통한 대사증후군 개선 효과 확인 연구에서는 검진 및 문진 내역

이 있는 연구에 참여한 국가 검진 대상자 중 사후관리 프로그램을 실제 3개월간 사용한 대상자를 선정하였다. 건강검진의 신체 측정 자료를 통해 체질량지수, 허리둘레, 수축기 혈압, 이완기혈압의 측정값을 이용하였으며, 건강검진의 혈액 검사 자료를 통해 공복혈당, 중성지방, 고밀도콜레스테롤의 결과를 이용하였다. 또한 생활습관에 대해 운동 횟수, 흡연력, 음주력, 수면 시간, 스트레스 정도와 비만, 혈압, 고지혈증, 당뇨 관련 영양 습관에 대하여 확인하였다.

건강in에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고에서는 총 5회의 전문가 초청 포럼과 최종 토론회를 하였고, 미래의 건강in의 방향으로 신뢰성 있는 대표적인 건강, 의료 플랫폼으로의 구축을 위한 제언과 함께 앱을 통한 개인 맞춤형서비스의 고도화에 관련한 논의를 하였다.

### 3. 연구 결과

간암질환 예측 연구의 일반검진 대상자는 최대 10년의 추적관찰 기간 중에 전체 대상자 중 1.72%인 150,922명에서 간암이 발생하였다. 남성에서 더 발생하며, 나이는 70대에서 가장 많이 발생하였다. 경제수준은 의료 급여와 의료보험 상위 4분위에서 높게 나타났다. BMI, 허리둘레가 증가한 경우 높았다. 당뇨, 고혈압이 있는 경우 간암발생이 높았으며 간암 있는 군에서 혈압, 혈당 높게 측정되었다. 간수치 중 AST, ALT, rGT 모두 간암군에서 높았으며 지방간을 예측하는 fatty liver index 역시 간암군에서 높았다. 콜레스테롤 지표는 HDL 포함하여 LDL, Triglyceride, cholesterol 모두 낮게 나타났다. B형 간염 여부에서 가장 큰 빈도차이를 보였으며, C형 간염, 기타 간염, 간경변, 나이 등에서 큰 빈도차이를 보였다.

생활습관의 경우 흡연력은 경우 간암이 더 많이 발생하였다. 음주에서는 저위험 음주력이 있는 경우는 상대적으로 낮았으나 그 외에는 음주량과 비례하는 경향을 보였다. 운동량은 큰 상관성을 보이지 않았다. 예측모델에서는 B형간염 진단 여부 및 나이에 따라 큰 격차를 보였으며, 고혈압과 운동량을 제외한 대부분의 변수들이 유의하였다.

일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측 모델에서는 나이가 가장 강력한 Hazard ratio를 나타냈다. 80세 이상은 HR이 5.361이었다. 흡연, 음주, 운동 등 생활습관이 악화될수록 HR를 증가하는 결과가 나타났다. 흡연하는 경우 HR는 1.157, 고위험 음주하는 경우 HR는 1.527으로 간암의 위험도가 올라갔다.

간암질환 예측 연구의 간암검진 대상자는 최대 10년의 추적관찰 기간 중에 전체 대상자 중 10.14%인 17,390명이 간암 발생간초음파검사 관찰조건에서 이형결절이 진단된 경우 가장 큰 빈도차이를 보였으나, HBV 42,077명 중 7,916명(18.81%), HCV 8,615명 중 1,637명(19%), 간경변 70,892명 중 8,863명(12.50%)에서 간암이 발생하였다. 나이는 60대에 가장 높은 발병을 보였다. 남성에서 간암 발생이 증가하였으며, 체중, 혈압 AST, ALT, rGT는 간암발생과 비례하여 증가하였다. 콜레스테롤 수치인 저밀도콜레스테롤, 중성지방, 콜레스테롤은 간암 군에서 더 낮게 나타났다. 고밀도 콜레스테롤 역시 낮은 경우 간암발생이 높았다. 생활습관에서는 고위험 음주 많이 상관성을 보였다. 이는 일반 검진자 특성과는 다른 경향을 보였다. 예측모델에서는 간초음파 관찰 소견상 이형결절이 가장 큰 격차를 보이는 것으로 나타났으며, 일반 검진의 결과와 유사하게 B형 간염, C형 간염, 간경변 등의 기저질환 진단여부 및 나이에 따라 차이가 나타났다. 간암검진 대상자 예측모델에서 추가된 FLI, AFP, 간초음파 관찰소견(거친에코상, 이형결절) 등의 변수에서 큰 차이가 나타났다, 그 외에 대부분의 변수들이 유의하였다.

대장암질환 예측 연구의 일반검진 대상자는 최대 10년의 추적관찰 기간 중에 전체 대상자 중 1.2%인 105,454명에서 대장암이 발생하였다. 70대에서 발생률이 가장 높았고 여성보다 남성에서 더 발생하였으며, 경제수준은 의료 급여에서 발생률이 가장 높게 나타났다. 고혈압, 당뇨병이 있는 경우 대장암 발생률이 높았다. 콜레스테롤 지표는 Total cholesterol, TG, LDL 모두 대장암 발생 군에서 높게 측정되었고, HDL은 대장암 발생군에서 낮게 측정되었으며, 고콜레스테롤인 경우 대장암 발생률이 높았다. 비만 1단계와 2단계에서 높은 발생률을 보였다. 생활습관의 경우 과거 흡연 경험이 있는 군에서 대장암이 더 많이 발생하였다. 음주에서는 고위험 음주력이 있는 경우 대장암 발생률이 높았으며, 운동량은 고활동군에서 대장암 발생률이 높았다. 그 외에 궤양성 대장염과 크론병을 진단 받은 경우 대장암 발생률이 높았다. 예측모델에서 나이에 따라 가장 큰 격차를 보였으며, 그 외 대부분의 변수들이 유의하였고 그 중 당뇨병, 음주에서 가장 큰 차이를 보였다.

대장암 질환 예측 연구의 대장암검진 대상자는 최대 10년의 추적관찰 기간 중에 전체 대상자 중 2.21%인 38,358명에서 대장암이 발생하였다. 80대에서 발생률이 가장 높았고 여성보다 남성에서 더 발생하였다. 경제수준은 의료 급여에서 발생률이 가장 높게 나타났다. 고혈압, 당뇨병이 있는 경우 대장암 발생률이 높았다. 콜레스테롤 지표는 Total cholesterol, HDL, LDL 모두 대장암 발생 군에서 낮게 측정되었고, TG는 대장암 발생군

에서 높게 측정되었으며, 고콜레스테롤인 경우 대장암 발생률이 높았다. 비만 1단계에서 가장 높은 발생률을 보였다. 생활습관의 경우 과거 흡연 경험이 있는 군에서 대장암이 더 많이 발생하였다. 음주에서는 고위험 음주력이 있는 경우 대장암 발생률이 높았으며, 운동량은 고활동군에서 대장암 발생률이 높았다. 그 외에 크론병을 진단 받은 경우와 분변잠혈검사 음성인 경우 대장암 발생률이 높았다. 예측모델에서 나이에 따라 가장 큰 격차를 보이는 것을 알 수 있으나 50세 이상의 대상에서는 일반 검진보다 그 격차가 줄어들었다. 그 외에 대부분의 변수들이 유의하였고 그 중 당뇨병, 음주, 분변잠혈검사에서 가장 큰 차이를 보였다.

대장암 질환 예측 연구의 분변잠혈검사 양성 대상자는 최대 10년의 추적관찰 기간 중에 전체 대상자 중 4.36%인 4,472명에서 대장암이 발생하였다. 80대에서 발생률이 가장 높았고 여성보다 남성에서 더 발생하였다. 경제수준은 의료 급여에서 발생률이 가장 높게 나타났다. 고혈압, 당뇨병이 있는 경우 대장암 발생률이 높았다. 콜레스테롤 지표는 Total cholesterol, HDL, LDL 모두 대장암 발생 군에서 낮게 측정되었고, TG는 대장암 발생군에서 높게 측정되었다. 생활습관의 경우 과거 흡연 경험이 있는 군에서 대장암이 더 많이 발생하였다. 음주에서는 고위험 음주력이 있는 경우 대장암 발생률이 높았으며, 운동량은 고활동군에서 대장암 발생률이 높았다. 그 외에 내시경 검사와 이중 조영검사 미수검자인 경우 대장암 발생률이 높았다. 예측모델에서 나이에 따라 가장 큰 격차를 보였으며, 그 외에 음주, 소득에서 가장 큰 차이를 보였다.

국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선효과 연구에서는 최종적으로 총 63명의 연구대상자의 결과를 분석하여 효과를 확인하였다.

대사증후군 개선효과의 측면으로는 사후관리 프로그램을 사용하기 전보다 후에 대사증후군 타입 비율이 3.2% 감소하였고, 대사나이 평균이 0.48세 감소하였으며 대사나이 타입 '나쁨' 비율이 17.5% 감소하였다.

생활습관 개선효과의 측면으로는 사후관리 프로그램을 사용하기 전보다 후에 운동 횟수 평균이 0.27회 증가하였고, 흡연자는 1명 감소하였다. 음주의 경우, 음주횟수 평균이 0.3회 감소하였고 음주량(잔) 평균이 1.17잔 감소하였다. 수면시간의 경우, '7-8시간' 수면 해당자가 10명 증가하였고 스트레스 점수의 평균이 0.01점 감소하였다. 비만, 혈압, 고지혈증, 당뇨 등 4가지 항목의 영양 습관이 좋지 않은 사람이 평균치로 약 2명 감소하였다. 또한, 사후관리 프로그램을 사용하기 전보다 후에 걸음수 평균이 323보 증가하였다.

건강iN은 국민건강 증진을 위하여 맞춤형 건강정보의 고도화로 계속 발전해 왔다.

대국민 교육 자료로는 질병관리본부에서 관리하고 있는 국가건강정보포털이 있고, 환자에 대한 기록으로는 각 의료기관의 정보와 함께 진료정보교류사업을 통한 EHR Data가 수집되고 있다. 국가에서 주도하고 있는 4차산업혁명위원회 디지털헬스특위에서는 예방·수요자 중심의 개인 의료데이터 연계·활용을 통한 국민건강증진 및 의료서비스 혁신을 목표로 의료데이터 이용 활성화로 나아가고 있다. 국가 PHR 사업 방향에 맞추어 건강iN의 방향을 설정해 가야하며, 건강 증진 및 질병예방을 위한 개인별 맞춤형 자료도 계속 고도화 해 나가야 한다.

#### 4. 결론 및 제언

간암질환 예측 연구에서는 일반검진대상자, 일반검진 대상자 중 간암위험도 높은 군 제외한 대상자, 간암검진 대상자 세 가지 군에 대해 분석을 하였다. 2018년까지의 최근 Data를 이용한 이번 연구는 기존 암 예측 연구와 비교한 결과 HR가 비슷하여 상대적으로 의미 있는 모형이었다고 할 수 있다. 추후 간암예측모형의 검증력을 높이기 위해서는 간암 발생과 관련될 수 있는 가족력 및 그 외 생략된 변수들을 도입하여 고도화된 모형을 구축할 필요가 있다. 본 연구의 한계점으로는 고위험군(간암검진 대상자)을 대상으로 예측모형을 만드는 것은 기존 HBV 항바이러스제 복용 여부, virus titer, HBeAg status 등 공단 자료로 보충하기는 어려운 자료의 획득의 문제점이 있다.

대장암질환 예측 연구에서는 일반검진, 대장암검진, 분변잠혈검사 양성 수검자에서 공통적으로 연령 증가에 따라, 남성인 경우, 고혈압, 당뇨병, 고콜레스테롤증에서 대장암 발생률과 위험도가 높은 결과를 보였다. 생활습관에서는 흡연자, 고위험 음주군에서 대장암 발생률과 위험도가 높은 것으로 나타났다. 이러한 위험군에 대하여 대장암 발생 예방을 위해 생활습관 관리가 필요하다. 대장암검진 대상자의 예측값이 낮게 분석되었는데, 이는 대장암 가족력이나 식이와 같은 요인은 대장암 발생 예측에 중요하지만 2009년 검진 자료에 관련 data가 없어 포함시키지 못 하였으며, 추후 국가검진에 사용하는 문진표에 보다 자세한 데이터 수집이 필요할 것으로 사료된다. 분변잠혈검사 양성 대상자에서 대장내시경이나 이중조영검사를 받지 않은 대상자에서 대장암 발생이 더욱 높은 빈도로 나타났으며, 추후 분변잠혈검사 양성인 수검자가 내시경 검사를 받는 비율을 높일 수 있는 방안이 필요하다.

사후관리 프로그램을 통한 대사증후군 개선효과 연구에서는 디지털헬스에서 전자기



기를 이용한 의료 접근에 대한 인식이 점차 보편화되는 시점에서 개인 맞춤형 DTx를 개인에게 적용하여 효과를 확인해 보았다. 약 3개월간 모집된 짧은 기간의 대상자 모집과 결과 분석의 한계로 분석 대상자의 50.8%인 63명의 적은 인원으로 분석이 진행되었으며 통계적 유의성을 입증하는데 한계점이 있었다. 또한 COVID-19 상황으로 인하여 실외 활동이나, 병원 방문 기피 현상 등으로 인해 모든 참여 대상자들의 개선 효과를 비교해 보는 데는 어려움이 있었다. 추후 연구 대상자 모집 기간 및 모집단 수를 늘려 통계적 유의성을 보완하는 연구 진행이 필요할 것으로 사료된다. 또한 검진 결과 제공과 함께 대사증후군 위험요인이 있는 대상자들에게 제공할 맞춤형 건강관리 프로그램의 지속적인 고도화가 필요하다.

국민건강보험공단의 건강iN은 평균적·보편적·대표적 PHR로 기능하도록 공공의 자료를 통합하여 나아가야 하며, 기업은 지속적·연계적인 PHR service를 창의적으로 제공하는 등 역할에 있어서 차이가 있다. 공단 data가 국민에게 제공되는 건강증진, 건강개선 서비스는 목표를 세분화해서 연령, 지역, 질병, risk factor 별로 개발이 필요할 것이며, 현재 시점에서 할 수 있는 서비스가 무엇인지 define 하는 것이 필요하다. 사용할 data에 대해서는 개인별 맞춤형 자료를 계속 고도화해야 한다. 국가검진 표준 연구 및 노화 관련 질병 발생 예측 등의 지속적인 개발이 필요하다. 기 개발된 여러 공공의 자료의 포괄적 활용과 함께 부족한 부분은 계속 개발해 나아가야 한다.

# 제 1 장

## 서 론

제1절 연구 배경 및 필요성	11
제2절 연구의 목적	12

---



# 제1장 서론

## 제1절 연구 배경 및 필요성

현재 세계적 연구의 트렌드는 일개 병원 데이터를 이용하여 개별화된 방법으로 연구를 진행하는 것이 아니라 표준화된 방법으로 다량의 정보를 수집하여 빅데이터를 통해 분석을 진행하는 것이다. 국민건강보험공단에서도 이러한 변화의 흐름에 맞추어 보험공단의 빅데이터를 이용하여 맞춤형 서비스를 건강iN(hi.nhis.or.kr)을 통해 구축해 다른 추가적 검사를 진행하지 않고 국가검진 데이터만으로 일반인들에게 질병예측에 관한 정보를 전달하여 국민 건강을 향상시키고자 노력해왔다.

본 연구는 개인건강기록 서비스(PHR)의 기초 연구로 시작하여 2014년, 2015년, 2016년, 2017년, 2018년, 2019년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발에 이은 후속 7차년도 연구이다. 올해 연구에서는 국민건강보험공단의 빅데이터실에서 보유하고 있는 맞춤형 DB를 통하여 수요자 중심의 맞춤형 서비스를 개발하는 것으로 간암, 대장암 예측 프로그램의 업데이트와 추가 개발을 진행하였다. 또한 국가검진대상자들의 사후관리 프로그램 참여를 통해 맞춤형 모바일 건강관리 프로그램의 효과 검증 및 생활습관과 건강상태 개선 효과를 확인하였고, 건강iN-전문가포럼을 통하여 앞으로 건강인의 나아갈 바를 논의해 보았다.

『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 메시지를 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 검증 한 바 있다. 『2017년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 위암, 폐암질환 예측모형 개발, 당뇨 질환 위험 예측 모델 고도화 및 건강인 웹페이지 탑재 알고리즘 개발, 심뇌혈관 질환 예측 모델 고도화, 각 생체지표의 연령별, 성별 국가 현황과 생활습관 항목과의 연관성 조사 및 건강노화나이 고도화 개발을 진행하였다. 『2018년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 국가적 극복 과제인 암을 중심으로 하여 위암 및 폐암의 예측모델을 개발 및 고도화하였고, 심뇌혈관의 가장 중요한 위험요인인 대사증후군의 현황과 대사증후군 나이를 연구하여 개인의 대사증후군 정도를 쉽게 계량화 하는 방안을 모색해 보았다. 또한 미래 극복과제인 치매에 대하여 산정특례를 이용하여 위험요인 현황을 알아보았다. 『2019년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 맞춤형 DB 및 암 검진 자료를 연계하여 질병 예측 프로그램을 개발하고, 대사 증후군의 표현형과 유전형의 상관성 분석을 통해 현재 건강인 이용자에게 개인별 맞춤 건강 정보를 제공하고 이를 통해 건강생활 습관을 유도하여 질병 발생의 예방과 삶의 질 향상에 도움이 되고자 하였다. 『2020년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』에서는 간암, 대장암의 일반대상자 및 암검진 대상자 각각에 대하여 위험요인을 분석해보고 예측모형을 개발하였다. 국가검진대상자 중에서 사후관리 프로그램 참여 동의를 대상으로, 디지털헬스를 이용한 맞춤형 모바일 건강관리 프로그램의 효과 검증과 함께 생활습관과 건강상태 개선 효과를 확인하고 이를 통해 건강생활 습관을 유도하여 질병 발생의 예방과 삶의 질 향상에 도움이 되고자 하였다.

건강in-전문가포럼에서는 의무기록에서의 EMR & EHR 데이터, PHR 의 현황과 방향, 진료정보교류 현황, 국민건강교육자료 등의 주제에 대하여 현황과 방향을 살펴보고 앞으로의 정보화 사회에서의 건강인의 방향을 논의하였다.

## 제2절 연구의 목적

본 연구는 『2014년 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발』 연구에 이은 후속 7차년도 연구로서, 맞춤형 DB를 통하여 질병 예측 모델의 고도화 및 개발, 국가검진대상자들의 사후관리 프로그램 참여를 통해 맞춤형 모바일 건강관리 프로그램의 효과 검증 및 생활습관과 건강상태 개선 효과 확인, 건강in-전문가포럼을 통해 이용자

의 건강, 질병정보 제공을 통하여 질병 발생의 예방과 질병 관리에 도움이 되고자 한다.

- 1) 간암 모델 고도화
- 2) 대장암 모델의 개발 : 일반검진 및 암검진 요인 분석
- 3) 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과
- 4) 건강in 에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고



# 제2장

## 이론적 고찰

제1절 간암 질환 예측	17
제2절 대장암 질환 예측	23
제3절 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과	25
제4절 건강iN에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고	27





# 제2장 이론적 고찰

## 제1절 간암 질환 예측

간암의 사망률은 2017년 국가 암 등록 통계자료에 따르면 남성의 경우 전체 암 사망률의 5위(9.4%), 여성의 경우 전체 암 사망률의 6위(3.6%)를 차지하고 있고, 특히 간암은 전체 연령군별 주요 사망률에서는 5위를 차지하고 있다. 간암은 전체 암 중 의료비 사용 면에서도 2위를 차지하고 있어 국가적 차원에서 집중해서 관리해야 할 암이다. <표 2-1>

<표 2-1> 성별 주요 암 발생 현황, 2017년도 국가암등록통계자료

(조발생률, 명/10만 명)

순위	남자					여자				
	암종 (16 순위)	발생자수	분율	조 발생률	표준화 발생률*	암종	발생자수	분율	조 발생률	표준화 발생률*
	모든 암	122,292	100.0	478.1	301.6	모든 암	109,963	100.0	428.6	278.7
	갑상선암 제외	116,257	-	454.5	280.7	갑상선암 제외	89,828	-	350.2	209.8
1	위	19,916	16.3	77.9	47.5	유방	22,300	20.3	86.9	63.0
2	폐	18,657	15.3	72.9	42.7	갑상선	20,135	18.3	78.5	68.9
3	대장	16,653	13.6	65.1	39.9	대장	11,458	10.4	44.7	23.0
4	전립선	12,797	10.5	50.0	29.0	위	9,769	8.9	38.1	21.1
5	간	11,500	9.4	45.0	27.6	폐	8,328	7.6	32.5	15.8
6	갑상선	6,035	4.9	23.6	20.8	간	3,905	3.6	15.2	7.4
7	췌장(10)	3,733	3.1	14.6	8.8	자궁경부	3,469	3.2	13.5	10.5
8	신장(9)	3,617	3.0	14.1	9.6	췌장	3,299	3.0	12.9	6.0
9	담낭 및 기타담도(7)	3,555	2.9	13.9	8.1	담낭 및 기타담도	3,291	3.0	12.8	5.5
10	방광(8)	3,525	2.9	13.8	8.2	자궁체부	2,986	2.7	11.6	8.1

\*연령표준화발생률: 우리나라 2000년 주민등록연앙인구를 표준인구로 사용  
출처: 2017년 국가암등록 통계자료(보건복지부, 중앙암등록본부, 국립암센터)

<표 2-2>연령 군별 주요 암발생률: 남녀전체, 2017, 국가암등록통계자료

(조발생률, 명/10만 명)

순위	0-14세	15-34세	35-64세	65세 이상
	모든 암 (14.1)	모든 암 (66.8)	모든 암 (470.1)	모든 암 (1542.2)
1	백혈병 (4.1)	갑상선 (31.9)	갑상선 (80.1)	폐 (261.9)
2	뇌 및 중추신경계 (2.1)	유방 (5.4)	유방 (74.3)	대장 (217.4)
3	비호지킨 림프종 (2.1)	백혈병 (3.0)	위 (58.9)	위 (215.0)
4	갑상선 (0.6)	대장 (2.5)	대장 (51.4)	전립선 (138.5)
5	난소 (0.4)	비호지킨 림프종 (2.5)	폐 (34.7)	간 (107.4)

연령표준화발생률: 우리나라 2000년 주민등록인구를 표준인구로 사용

간암은 증상을 유발하지 않는 경우가 많아 발견 시 치료하기 어려운 병기에 발견되는 경우가 많다. 간암이 초기에 발견되는 경우 간절제술, 간이식, 고주파열치료술 등 근치적인 치료를 받을 수 있으나, 진행성 간암으로 진단을 받게 되는 경우 전신항암요법 등 비근치적 치료를 받게 되는 경우가 많다. 따라서 간세포암종의 고위험군을 선정하고, 이를 조기진단을 위한 선별 효과적인 치료와 생존기간의 향상을 위하여 중요하다. 그래서 각 국가에서는 간세포 암 종의 고 위험군을 선정하기 위해, 초음파검사 및 α-FP등의 선별검사를 이용한 조기진단을 하고 있다.

세계 각국에서는 간암을 조기 진단하고 예측하기 위해 간암예측모형들을 개발하고 있다. 태국과 홍콩에서는 18세 이상의 health care system 자료 간암예측모형을 개발하였다. 이 예측모형에서는 항바이러스제 치료를 받은 B형간염 환자를 대상으로 HCC 발생을 예측하는 CAMD score를 계산하였다. 그 결과는 40세 이하이며 간경화가 동반된 경우 10 점, 60세 이상 군에서 10점으로 가장 높게 발생위험이 측정되었다. <표 2-3>

<표 2-3> multi variable cox model for HCC occurrence and CAMD score

Variables	Adjusted hazard ratio	CAMD score
<b>Cirrhosis</b>		
No cirrhosis	Reference	0
Cirrhosis with age <40 years	18.8 (95% CI 9.2–38.7)	10
Cirrhosis with age ≥40 years	4.6 (95% CI 3.8–5.6)	6
<b>Age (years)</b>		
<40	Reference	0
40–49	4.5 (95% CI 2.4–8.5)	5
50–59	9.0 (95% CI 4.8–16.8)	8
≥60	15.9 (95% CI 8.5–29.7)	10
<b>Male sex</b>		
Female sex	Reference	0
Male sex	1.8 (95% CI 1.4–2.2)	2
<b>Diabetes mellitus</b>		
Not diabetic	Reference	0
Diabetic	1.3 (95% CI 1.1–1.6)	1

캐나다에서는 Liver Centre에서 간경화로 예상되는 인원 대상으로 cohort연구를 진행하였다. 이 코호트를 바탕으로 2000년 1월 1일~2009년 12월 31일간 총 10년간 간암을 발생률을 예측하는 모형을 구축하였다. Univariable analysis 결과 Harzards ratio는 HCV 6.74, HBV 6.50 steatohepatitis 4.14로였다. multivariable analysis 결과 Harzards ratio는 HCV 3.83, 60세 초과 3.77, HBV 3.66 순서로 높았다. <표 2-4 >

<표 2-4> Factors associated with HCC by cox proportional hazards regression

Covariate	Univariable HR (95% CI)	p value	Multivariable HR (95% CI)	p value
<b>Etiology</b>				
Autoimmune	1		1	
HCV-SVR	1.74 (0.65–4.67)	0.27	1.04 (0.38–2.83)	0.94
Other	2.46 (0.95–6.34)	0.06	1.65 (0.63–4.27)	0.31
Steatohepatitis	4.14 (1.80–9.51)	0.001	2.12 (0.91–4.92)	0.08
HCV	6.74 (3.13–14.50)	<0.001	3.83 (1.76–8.33)	0.001
HBV	6.50 (2.99–14.16)	<0.001	3.66 (1.66–8.07)	0.001
Age	1.03 (1.02–1.05)	<0.001	1.04 (1.03–1.06)	<0.001
<b>Age category</b>				
<45	1		1	
45–60	1.91 (1.27–2.87)	0.002	1.94 (1.28–2.93)	0.002
>60	2.94 (1.94–4.46)	<0.001	3.77 (2.45–5.80)	<0.001
<b>Male Sex</b>				
Platelets ( $\times 10^9/L$ )	2.84 (2.04–3.94)	<0.001	2.92 (2.07–4.12)	<0.001
Platelets ( $\times 10^9/L$ )	0.99 (0.99–0.99)	<0.001	0.99 (0.99–1.00)	<0.001
<b>Platelet Category (<math>\times 10^9/L</math>)</b>				
<80	1		1	
80–139	0.78 (0.57–1.07)	0.12	0.78 (0.57–1.07)	0.13
140–200	0.40 (0.27–0.59)	<0.001	0.40 (0.27–0.60)	<0.001
>200	0.25 (0.16–0.41)	<0.001	0.31 (0.19–0.50)	<0.001
<b>APRI</b>				
<1.5	1			NS
1.5–4	1.69 (1.26–2.28)	<0.001		
>4	1.57 (1.09–2.26)	0.01		
<b>FIB-4</b>				
0–1.5	1			NS
1.5–5.5	2.55 (1.44–4.53)	0.001		
>5.5	4.10 (2.29–7.34)	<0.001		
MELD >8	1.61 (1.23–2.11)	0.001		NS

AST, aspartate aminotransferase; APRI, AST to platelet ratio index; CI, confidence interval; FIB-4, Fibrosis-4 score; HBV, Hepatitis B virus; HCV, Hepatitis C virus; HR, Hazard Ratio; MELD, model for end-stage liver disease; SVR, sustained virologic response.

미국에서는 National Liver Transplantation Center 대기자 자료를 이용하여 34,932명  
 에 대한 코호트를 구축하였다. 이 코호트를 분석하여 2002년~2011년 10년간 간암이  
 발생을 예측하는 모형을 구축하였다. 여러 변수(age, sex, race, etiology of cirrhosis,  
 DM, BMI, Hx of TIPS, Hx of variceal bleeding, bilirubin, INR, Cr)를 활용하여 Cox  
 proportional hazard regression로 ADDRESS-HCC 예측모델 개발하였다. 그 결과 HR은  
 viral 4.72, alcohol/metabolic liver disease 2.28, 남성 2.03, 당뇨 1.31, INR 1.36,  
 CTP(Child turcotte-pugh) 1.17, MELD(model of end-stage liver disease score) 1.02,  
 순서로 높았다. <표 2-5>

<표 2-5> Cox models evaluating risk factors for the ADDRESS-HCC model

Variable	Univariate Model HR (95% CI) P		Multivariate Model HR (95% CI) P	
Age (per y increase)	1.04 (1.03-1.04)	<.001	1.05 (1.04-1.06)	<.001
Etiology of cirrhosis <sup>a</sup>				
Autosomal	Reference	-	Reference	-
Alcohol/metabolic	2.28 (1.69-3.07)	<.001	1.43 (1.04-1.97)	.030
Viral	4.72 (3.55-6.26)	<.001	3.38 (2.50-4.55)	<.001
Diabetes present	1.31 (1.11-1.54)	.001	1.25 (1.06-1.48)	.008
Male sex	2.03 (1.76-2.33)	<.001	1.99 (1.71-2.30)	<.001
CTP score (per unit increase)	1.17 (1.13-1.21)	<.001	1.17 (1.13-1.21)	<.001
Nonwhite race	1.26 (1.12-1.43)	<.001	1.20 (1.06-1.37)	.005
MELD score (per unit increase)	1.02 (1.01-1.04)	.007	-	-
Body mass index (per 5-unit increase)	1.01 (0.96-1.07)	.681	-	-
Presence of TIPS	0.88 (0.69-1.12)	.315	-	-
Previous variceal hemorrhage	0.93 (0.67-1.29)	.643	-	-
Bilirubin (per unit increase)	1.03 (1.01-1.05)	<.001	-	-
INR (per unit increase)	1.36 (1.22-1.52)	<.001	-	-
Creatinine (log)	1.16 (1.00-1.33)	.040	-	-

Abbreviations: 95% CI, 95% confidence interval; ADDRESS, age, diabetes, race, etiology cirrhosis, sex, and severity; CTP, Child-Turcotte-Pugh; HCC, hepatocellular carcinoma; HR, hazards ratio; INR, international normalized ratio; MELD, model of end-stage liver disease; TIPS, transjugular intrahepatic portosystemic shunt.

국내에서는 1990년~1998년까지 만성간질환을 가진 환자 4339명을 추적 관찰하여  
 간암 발생의 위험인자에 따른 간암 발생 예측 모형 등을 개발하였다. 그 결과 간경화HR  
 5.6, 40세 이상 HR 3.7, HCV의 HR 3.5, HBV의 HR 2.2 이었다.

<표 2-6> significant risk factor for HCC

Factors	Univariate	Multivariate	
	<i>p</i> -value	OR (95%CI)	<i>p</i> -value
Status of disease			
Liver cirrhosis (n=335)	0.000**	5.6 (2.51- 14.03)	0.000**
Chronic hepatitis (n=540)	0.001**	2.1 (1.24- 6.80)	0.049*
Carrier (n=119)		1.0	
Type of hepatitis			
HCV (n=121)	0.000**	3.5 (1.33- 6.20)	0.016*
HBV (n=781)	0.001**	2.2 (1.00- 3.55)	0.089
NonBNonC (n=92)		1.0	
Age (years)			
> 40 (n=798)	0.000**	3.7 (1.70- 4.76)	0.014*
≤ 40 (n=742)		1.0	
Liver parenchymal echogenic pattern			
Severe (n=247)	0.001**	1.8 (0.95- 2.65)	0.073
Normal/Mild (n=747)		1.0	
Initial ALT level at enrollment			
> 40 IU/L (n=552)	0.003**	1.3 (0.99- 2.37)	0.261
≤ 40 IU/L (n=442)		1.0	
Initial serum αFP level at enrollment			
> 20 ng/mL (n=191)	0.001**	2.3 (0.92- 2.42)	0.001**
≤ 20 ng/mL (n=803)		1.0	
Drinking			
Heavy (n=149)	0.005**	1.8 (0.83- 2.40)	0.073
Non/ Social (n=543)		1.0	
Unknown history (n=302)		1.2 (0.57- 2.14)	
Sex			
Male (n=683)	0.012*	1.4 (0.71- 1.68)	0.686
Female (n=311)		1.0	

\*, *p*<0.05; \*\*, *p*<0.01; OR, odds ratio; CI, confidence interval.

각 연구에서 예측 모형에 구축에 사용된 변수들 종합하면 아래와 같다.

<표 2-7> 기존 간암예측모형에서 활용된 위험요인들

	Kim et al (2018, 한국)	Hou et al (2018, 미국, 중국)	Yano et al. (2011, 미국)	Lognoui et al (2018)	Tomoko et al (2017, 일본)	Fleming et al (2014, 미국)	Gural et al. (2018, 캐나다)	Takehiro et al (2012, 일본)	Wen et al (2012, 미국)	Yuna et al (2015, 미국)	Li et al (2018, 미국)
population	HBV	HBV	HBV	HCV	CLD	cirrhosis	cirrhosis	general	general	general	DM
Age											
Sex											
Race											
BMI											
Smoking											
Alcohol											
Physical activity											
FHx											
Etiology	HBV		HBV or HBsAg								
	HCV										
	Autoimmune										
	Other (alcoholic)					metabolic	steatosis		CU		
Esophageal varix											
Cirrhosis											
DM											
AST											
ALT											
Platelet											
Glucose											
Albumin											
AFP											
간기능 score											
DM med											
Dyslipidemia med											
기타						INR, Cr, bilirubin					

각 국가별 예측모형에서 간암을 일으키는 인자들의 영향성은 차이를 보였다. 특히 동아시아 국가 내에서도 중국의 경우 간암 발생의 80%가 B형 간염 바이러스 감염과 관련성이 높으며, 일본은 간암 발생의 75%가 C형 간염 바이러스 감염과 관련이 있는 것으로 알려져 있다. 따라서 우리나라 자료로 간암을 예측할 수 있는 최신의 연구가

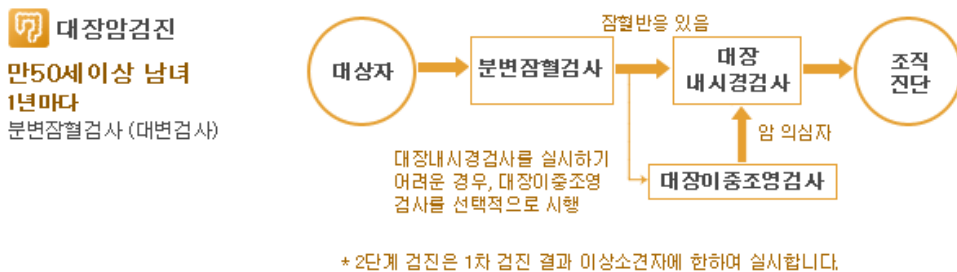
필요하겠다.

기존 예측모형들 중 국민전체 인구를 반영하는 자료로 한 연구는 드물었다. 하지만 간암이 없는 대상자 군에서도 간암예측을 하게 된다면, 국민전체의 간암 사망률을 낮추는데 기여하게 될 것이다. 이에 세계적으로 보험 가입률이 높은 국민건강보험공단의 자료를 활용하여 국민전체를 대상으로 한 간암예측 모형을 구축 하고자 한다. 또한 허리둘레, 체질량지수(BMI), 트리글리세라이드, 감마글루타밀전달효소(GGT)를 기반으로 한 알고리즘인 지방간지수(FLI)는 서양에서 지방간을 검출하기 위해 처음 개발됐다. 본 연구는 국민건강보험공단 자료로 간암 예측에 사용되는 수치인 Fatty Liver Index의 간암검진 활용성을 본다.

## 제2절 대장암 질환 예측

2017년도 국가 암 등록 통계자료에 의하면 10만 명 당 남성의 경우 16,653명 (13.6%), 여성의 경우 11,458명 (10.4%)로 남녀 각각 세 번째로 대장암이 많이 발생하는 것으로 조사되었다<표 2-1>. 연령군 별 주요 암발생률을 남녀 전체로 본 경우에도 15-64세에서 4위, 65세 이상에서도 2위를 차지할 정도로 발생률이 높은 암에 속해 더욱 집중해서 관리해야 할 암으로 판단된다<표 2-2>.

우리나라에서는 국가 암 검진 지원 사업을 통해 만 50세 이상 남녀를 대상으로 1년마다 분변잠혈검사를 실시하여 잠혈 반응이 있는 경우 대장내시경 검사 또는 대장이중조영 검사를 실시하는 대장암검진 사업을 수행하고 있다.



[그림 2-1] 국가 암검진 지원사업



대장암 발생 예측에 대한 <표 2-8>과 같이 여러 예측 모형들이 존재하며 모형 적합도는 70% 내외이다. 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 있다. 이 외에도 운동량, 식습관, 당뇨 등의 요소가 포함된 모델들이 있었다.

<표 2-8> 대장암 예측모형별 위험요인 비교분석

	Cai et al, 2012	Yeoh et al, 2011	Ma et al, 2010	Wei et al, 2009	Freedman et al and Park et al, 2009	Kastrinos et al, 2009	Driver et al, 2007	Lin et al, 2006	Imperiale et al, 2003	Levitzy et al, 2011	Korean
Age											남·여
Gender											남·여
Smoking											남·여
DM											
Green vegetables											
Pickled food											
Fried food											
FDR with CRC						CRC or Lynch		adenoma or CRC			
SDR with CRC											
BMI											남
Alcohol											남 frequency & amount
physical activity					leisure time activity (men only), leisure exercise time(women only)						
Meat consumption	white meat			red or processed meat							여
Folate											
Sigmoidoscopy results											
Colonoscopy results											
History of polyps											
Relative with CRC						more than 3 relatives					
Aspirin/NSAID Use											
Estrogen status (women only)											
CRC or polyps diagnosed before age 50											
Height											여
Fasting glucose											남
T.cholesterol											남
Family Hx of cancer											암

국내에서는 Shin et al. 이 2014년에 5년간 암이 발생할 위험을 예측하는 모형을 남녀에서 개별적으로 개발하였다. 30세에서 80세 사이의 인구를 대상으로 연구하였으며 1996년에서 1997년에 건강검진을 받은 846,559명의 남성과 479,449의 여성에서 향후 colorectal cancer의 발병 여부를 Korean Central Cancer Registry (KCCR) database 등록 자료로 확인하였다. 이 인구에서 모형을 개발하였으며 1998년과 1999년 사이에 건강검진을 받은 547,874명의 남성과 415,875명의 여성에서 모형의 적합도를 평가하였다. 이 연구에서 proximal/distal colon cancer 및 rectal cancer에 대하여 개별적 모형이

개발되어 있다. 또한 남성과 여성에서 colorectal ca, Lt/Rt colon ca, colon ca, rectal ca 등의 5가지 모형을 각각 개발하였으며 C-statistic으로 평가한 모형의 예측력은 0.69에서 0.78사이의 값으로 확인되었다. 남성에서는 모형에 의해 예측된 발병률과 실제결과가 매우 비슷하게 나타났으나 여성에서는 유의한 예측력을 보이지 않았다. <표 2-9>

<표 2-9> C statistic and Hosmer-Lemeshow type chi-square test for colorectal cancer risk prediction model for development set(8-year risk) and validation set(5-year risk)

Models	Colorectum	Right colon	Left colon	Colon	Rectum
<b>Male</b>					
Development set					
C (95% CI)	0.762 (0.755, 0.769)	0.740 (0.721, 0.759)	0.786 (0.772, 0.799)	0.767 (0.756, 0.778)	0.753 (0.743, 0.763)
Chi-square value (p-value)	14.567 (0.1035)	7.162 (0.6203)	8.567 (0.4782)	11.871 (0.2207)	10.788 (0.2906)
Validation set					
C (95% CI)	0.779 (0.768, 0.789)	0.762 (0.734, 0.789)	0.779 (0.758, 0.801)	0.770 (0.753, 0.787)	0.779 (0.766, 0.793)
Chi-square value (p-value)	31.383 (0.0003)	24.540 (0.0035)	8.189 (0.515)	18.575 (0.029)	30.970 (0.0003)
<b>Female</b>					
Development set					
C (95% CI)	0.706 (0.695, 0.718)	0.745 (0.722, 0.768)	0.678 (0.652, 0.704)	0.711 (0.693, 0.728)	0.698 (0.682, 0.714)
Chi-square value (p-value)	8.219 (0.5123)	6.720 (0.6663)	5.936 (0.7463)	6.054 (0.7345)	15.246 (0.0844)
Validation set					
C (95% CI)	0.726 (0.712, 0.741)	0.763 (0.736, 0.791)	0.690 (0.659, 0.721)	0.723 (0.702, 0.743)	0.716 (0.696, 0.737)
Chi-square value (p-value)	13.129 (0.1569)	4.597 (0.8679)	10.415 (0.3180)	8.421 (0.4924)	12.279 (0.1980)

doi:10.1371/journal.pone.0088079.t003

### 제3절 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과

의료 패러다임의 전환을 디지털 변환(digital transformation)으로 대응하고자 하는 스마트 헬스케어(smart healthcare)는 ‘정보통신 및 바이오, 나노 등의 기술을 활용하여 언제 어디서나 개인 맞춤형 건강관리 및 의료서비스 등을 제공하는 융합산업’으로 받아들여지고 있다. 이는 병원의 진료·의료 정보를 활용하는 서비스에 초점을 맞춘 e-헬스, 일상생활의 건강관리와 질병예방을 중심으로 하는 u-헬스, ICT 활용에 초점을 맞춘 디지털 헬스케어 등 다양한 모습으로 전개되며 진화하고 있다.

스마트 헬스케어가 주목받는 핵심 이유는 그 지향점이 ‘예측적이고 예방적이며 개인화되고 참여적인’, 이른바 P4(predictive, personalized, preventive, participatory)를 겨냥함에 있다. 즉 ‘시설중심에서 거주지로, 질병 치료중심에서 예방으로’라는 의료패러다임

전환의 실제적 실천 수단이 될 수 있을 뿐 아니라 제조업체, 병원 및 환자를 비롯한 기존 업계 종사자와 이해 관계자들 간의 관계를 변화시켜 기존 의료산업에서의 순차적 가치 사슬을 벗어나 플랫폼이라는 초연결 중심으로 재편될 수 있음을 의미한다.

특히 최근 1~2년 사이에 디지털 헬스케어 업계에서 떠오르는 주제가 디지털 신약(Digital Therapeutics, DTx)이다. DTx는 의학적 장애 또는 질병을 예방, 관리 또는 치료하기 위한 소프트웨어를 기반으로 하는 치료를 지칭하며, 소프트웨어 자체가 의약품과 같은 효능을 갖고 환자에게 작용하는 것이 궁극적 목표이다. 또한 의약품, 의료기기 등의 전통적인 치료법과 병행하거나 독립적으로 사용될 수 있으며, 특히 우울증 및 불면증과 같은 질환을 치료하는 전통적인 약물 치료를 대체 혹은 보조하거나, 금연과 같은 행동 교정에서 기존 치료법보다 효율적이고 효과적인 대안으로 발전할 수 있는 잠재력을 가진다.

현재 우리나라는 남녀노소를 불문하고 스마트폰 및 컴퓨터 사용에 익숙한 인구의 성장과 함께 5G와 같은 초고속 인터넷의 발달로 일상에서의 스마트폰 활용이 당연시되었다. 또한 원격진료 및 디지털헬스 등을 통해 전자기기를 이용한 의료에 대한 인식이 점차 보편화되고 있어, DTx는 장소와 시간에 구애받지 않고 지속적인 치료 및 환자에 대한 모니터링이 가능하다는 점이 큰 장점으로 작용한다.

전 세계는 장기화되는 COVID-19로 인해, 비대면 행위에 주목하게 되었고 특히 healthcare 분야에서 비대면 의료행위에 많은 관심과 투자를 하고 있는 실정이다.

이런 의료패러다임의 변화에 발맞추어 이번 연구는 비대면 시대에 개인 맞춤형 DTx를 개인에게 적용하였을 때의 효과를 확인해 보고자 한다. 특히 생활습관 코칭 및 개선을 통한 연구 참여 전후 비교를 해보고자 하는 일종의 pilot study를 진행하였다.

본 연구는 두 가지 연구 목표를 설정하여 진행 하였다.

첫째, 모바일을 활용한 개인 맞춤형 건강관리 서비스가 대사증후군 예방 및 개선에 미치는 영향을 파악하기 위해 국가검진대상자들의 참여를 유도하여 대사증후군 요인별로 제공되는 교육 자료의 분석 및 효과를 확인하고자 한다. 둘째, 개인 맞춤형 건강관리 서비스 제공을 통해 생활습관 및 건강상태의 개선 효과가 있는지 확인하고자 한다. 이는 만성질환 유병률을 낮춤으로써 궁극적으로는 개인의 사회·경제적 부담을 증가시키고 있는 의료비 절감의 효과가 있다.

본 연구에서는 기존에 제공되고 있는 건강관리 서비스를 언택트 시대에 맞춰 시·공간적 한계를 극복하고, 맞춤형 건강정보를 제공받을 수 있는 ICT 융합기술을 적용한 공공형

건강관리 서비스로 개인 맞춤형 건강관리를 통해 자가 건강관리 능력을 향상시켜 건강격차 해소 및 국민 건강증진에 기여하고자 한다.

## **제4절 건강iN에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고**

지금까지 7년차의 연구를 통하여, 건강iN에서의 다양한 건강·질병정보와 함께 검진자료를 이용한 다양한 맞춤형 건강정보 콘텐츠를 연구 지원해 왔다. 새로운 4차 산업혁명의 시대를 맞이하여 전문가들의 포럼을 통하여 앞으로의 방향을 논의하였다.



# 제3장

## 연구 내용 및 방법

제1절 연구 내용	31
제2절 세부 과제별 연구 방법	33

---



# 제3장 연구 내용 및 방법

## 제1절 연구내용

### 1. 자료 개요 - 국민건강보험 청구 DB

간암 및 대장암 연구에서 사용한 자료는 모두 맞춤형 자료이다. 맞춤형 자료란 국민건강보험공단이 수집, 보유, 관리하는 모든 수검자의 건강보험 및 장기요양보험 자료에서 연구에 맞는 변수를 요청하여 받은 DB이다. DB 항목 중 자격 및 보험료, 사망, 명세서(20t), 진료내역(T30), 처방교부상세(T60), 일반검진 문진표 및 결과, 암검진 문진표상 동반질환/과거력 등을 대상으로 하였다.

<표 3-1> 맞춤형 DB 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 변수로 구성</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·보건기관(T1), 치과·한방(T2), 약국(T3)자료에 대한 명세서(20t), 진료내역(30t), 상병내역(40t), 처방전교부상세내역(60t)의 10개 세부 DB로 구성</li> <li>변수 : 명세서 공통, 진료, 상병, 처방 관련 등의 변수로 구성</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, 2009~2018년 건강검진DB 별도 구성 ... 검진제도 개편(2009년)으로 주요 검진 및 문진항목 변경</li> </ul>
요양기관DB	<ul style="list-style-type: none"> <li>내용 : 요양기관의 종별, 설립구분별, 지역(시도)별 현황 및 시설, 장비, 인력 관련 자료</li> </ul>



## 2. 간암 질환 예측 연구

일반검진 수검자 대상으로 간암과 나이, 성별, 수입, 비만도, 허리둘레, 혈압, 당뇨, 고혈압, 사망률, 혈당, AST, ALT, 콜레스테롤, 생활 습관, HBV, HCV와의 관련성을 보았다. 일반검진 수검자 자료로 간암예측모형을 구축하고, 이 모형을 ROC curve를 구축하여 검증력을 판별한다. 특히 일반 검진 대상자 중 간암검진자의 위험군으로 선정한 HBV, HCV,간경화를 제거한 군을 기준으로 예측모형을 개발하였다. 이 모형을 ROC curve를 구축하여 검증력을 판별한다.

간암검진 수검자를 대상으로 나이, 성별, 소득, 고혈압, 당뇨병, 허리둘레, 콜레스테롤, 생활습관, HBV, HCV와 간암의 상관성을 본다. 특히 간암 건진대상자(B형 간염, C형 간염 보균자와 간경화 환자)들에게 6개월 마다 AFP 혈청 검사와 복부 초음파를 제공한다. 이 연구에서는 간암검진의 혈청 AFP,초음파 결과를 활용하여 간암과 상관성을 분석한다. 간암검진 수검자 자료로 간암예측모형을 구축하고, 이 모형을 ROC curve를 구축하여 검증력을 판별한다. 또한 국민건강보험공단 자료로 간암 예측에 사용되는 수치인 Fatty Liver Index의 간암검진 활용성을 본다.

## 3. 대장암 질환 예측 연구

일반검진 수검자를 대상으로 대장암과 나이, 성별, 소득, 비만도, 허리둘레, 혈압, 당뇨, 콜레스테롤, 생활 습관(흡연, 음주, 운동), 궤양성 대장염, 크론병과의 관련성을 분석하였다. 일반검진 대상자의 대장암 발생 예측모형을 구축하고, 이 모형을 ROC curve를 이용하여 모델 적합도를 평가하였다.

대장암 검진 수검자를 대상으로 대장암과 나이, 성별, 소득, 비만도, 허리둘레, 혈압, 당뇨, 콜레스테롤, 생활 습관(흡연, 음주, 운동), 궤양성 대장염, 크론병, 분변잠혈검사와의 관련성을 분석하였다. 대장암검진 대상자의 대장암 발생 예측모형을 구축하고, 이 모형을 ROC curve를 이용하여 모델 적합도를 평가하였다.

또한 대장암 검진 수검자 중 분변잠혈검사 양성자를 대상으로 대장암과 나이, 성별, 소득, 비만도, 허리둘레, 혈압, 당뇨, 콜레스테롤, 생활 습관(흡연, 음주, 운동), 궤양성 대장염, 크론병, 내시경 여부, 이중조영검사 여부와의 관련성을 분석하였다. 분변잠혈검사 양성자의 대장암 발생 예측모형을 구축하고, 이 모형을 ROC curve를 이용하여 모델 적합도를 평가하였다.

#### 4. 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과

본 연구는 국가검진대상자들의 사후관리 프로그램 참여를 통해 맞춤형 모바일 건강관리 프로그램의 효과를 검증하였다. 대사증후군 예방 및 개선에 미치는 영향과 생활습관 및 건강상태의 개선 효과를 Paired T-test와 Chi-square Test를 통해 확인하였다.

4. 건강in-전문가포럼에서는 의무기록에서의 EMR & EHR 데이터, PHR 의 현황과 방향, 진료정보교류 현황, 국민건강교육자료의 현황과 방향 등을 주제로 전문가 발표와 함께 최종 토론회를 5차례에 걸쳐서 시행하였다.

## 제2절 세부 과제별 연구 방법

### 1. 간암 질환 예측 연구

#### (1) 일반검진자 대상 간암 질환 예측을 위한 데이터셋 구축

2009년 국가일반검진 대상자 전수를 대상으로 한다. 결측치 및 검진수치 이상치 상위1%,하위 1%를 제외 한 후 총 8,787,095명을 대상으로 하였다. 2004년에서 2008년 간 간암 진단코드를 받은 대상자는 제외하고, 2018년까지 총 10년간의 간암 발생여부를 조사한다. 간암의 발생은 간암 주상병코드(C22)를 기준으로 한다.

#### (2) 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 질환 예측을 위한 데이터셋 구축

일반검진 대상자 총 8,787,095명 중 간암검진자의 위험군으로 선정한 HBV, HCV, 기타간염, 간경화를 제거한 인원수를 대상으로 데이터를 구축한다. 2004년에서 2008년 간 간암 진단코드를 받은 대상자는 제외하고, 2018년까지 총 10년간의 간암 발생여부를 조사한다. 간암의 발생은 간암 주상병코드(C22)를 기준으로 한다.



#### (4) 간암 예측 모델 개발에 포함될 변수 선정

앞서 살펴본 해외 연구 사례에 근거한 위험인자들은 다음과 같았다. 일반 특성으로 나이, 성별, 체질량 지수가 포함되었고 혈청 검사인 AST, ALT, 혈소판, 알부민, 원인 질환으로 HBV, HCV, CLD, autoimmune hepatitis, 알콜성 간염, 기타 대사질환에 의한 간염이 포함되었다. 또한 동반질환인 간경화, 식도정맥류, 당뇨, 고지혈증 진단 그 외 등이 해외 연구사례의 위험요인이었다. <표 2-6>.

##### ① 일반검진자 대상 위험인자 선정

과거 예측모형에서 사용된 위험요소들 중 국민건강보험 데이터에 존재하는 변수들을 본 연구에 포함시켰다. 최종적으로 이 연구에서는 나이, 성별, 수입, 체질량지수(kg/m<sup>2</sup>), 혈압(mmHg), 허리둘레(cm), AST(U/L), ALT(U/L), 콜레스테롤(mg/dL), 당뇨 유무, 고혈압 유무, 그리고 흡연, 음주, 운동 등이 위험요인으로 선정되었다. 나이는 40세 미만, 40-50, 50-60, 60-70, 70-80, 80세 이상으로 구분하였다. 소득은 의료급여층, 일반보험 대상자를 4분위하여 구분하였다. 체질량지수는 18.5 미만(저체중), 18.5~22.9, 23~24.9(과체중), 25~29.9(비만), 30이상(고도비만)으로 분류하였다. 당뇨병은 공복혈당 126 이상 또는 설문지에 의사에게 당뇨병 진단받은 경우로 정의하였다. 고혈압은 수축기 140이상 또는 이완기 90이상 또는 설문지상 고혈압을 진단받은 경우로 정의하였다. 그 외 간암과 관련된 동반질환으로는 간경변, B형 간염, C형 간염, 기타 간염을 포함시켰다.

생활습관은 다음과 같이 분류하였다. 흡연의 경우 비흡연, 과거흡연, 현재흡연 세 가지로 분류한다. 음주의 경우 WHO 권고 기준에 따라 평균 하루 섭취 알코올 용량을 분류한다. 남자의 경우  $0 < \text{저위험} < 40\text{g}$ ,  $40 \leq \text{중위험} < 60\text{g}$ ,  $60\text{g} \leq \text{고위험}$ 으로 분류하였다. 여자의 경우  $0 < \text{저위험} < 20\text{g}$ ,  $20 \leq \text{중위험} < 40\text{g}$ ,  $40\text{g} \leq \text{고위험}$ 으로 분류한다. 평균 하루 섭취 알코올 용량은  $A \times 50\text{ml}(\text{소주한잔 용량}) \times 0.22(\text{소주도수}) \times 0.8(\text{알콜 함량비}) \times B/7$ 으로 계산한다. (A: 하루 섭취하는 잔수, B: 1주일에 마시는 횟수). 운동의 경우 WHO에서 제공하는 IPAQ-K 기준에 의해 분류한다.  $0 \leq \text{저강도활동} < 599 \text{ MET} \times \text{min}$ ,  $600 \leq \text{중등도활동} < 2999 \text{ MET} \times \text{min}$ ,  $3000 \text{ MET} \times \text{min} < \text{고활동군}$ 으로 분류한다. 운동은 설문지에 근거하여 격렬한 신체활동을 한 시간 이상 한 경우 주당 횟수  $\times 8.0 \text{ MET} \times 60\text{분}$ 으로 계산하고, 중등강도 신체활동을 한 시간 이상 한 경우 주당 횟수  $\times 4.0 \text{ MET} \times 60\text{분}$ 으로 계산, 저강도활동을 30분이상한 경우 걷기 주당 횟수  $\times 3.3 \text{ MET} \times 30\text{분}$ 으로 계산한다.

## ② 간암 검진자 대상 위험인자 선정

일반검진자에게 사용한 위험인자를 모두 포함시켰다. 추가로 FLI(Fatty Liver Index), 국가에서 6개월마다 제공하는 AFP 혈청검사, 복부초음파 검사 결과를 위험인자로 추가했다. AFP는 정량 검사인 경우 8.8ng/mL 이상인 경우를 AFP 양성 결과로 정의하였고, 그 외 정성검사를 활용하였다. 간암 초음파 결과에서 간암의 심 소견은 제거하였다. 간초음파 결과지의 경우 3set로 분류되어있는데, 이 3건의 관찰소견 중 상대적으로 위험한 질환을 남기는 것으로 재분류하여 하나의 결과를 만들고, 이 결과를 거친 에코상, 간경변, 이형결절, 양성질환, 기타 여부의 개별 변수로 활용하여 예측모형에 도입하였다.

## (5) 분석 방법

암 발생에 따른 대상자의 기본적 특성 차이를 보기 위하여 Chi-square test 및 T-test 분석을 진행하였고, 암 경험이 없는 대상자에서 대장암이 발생하기 까지 소요 되는 시간과 발생 여부를 이용하여 건강검진정보와 대장암과의 연관성을 알아보기 위해 예측모형을 구축하였다. 전체 대상자를 70%의 Train set, 30%의 test set으로 나누어 분석하였다. Train set에서 Cox-proportional hazards model, Logistic regression 을 통하여 예측 모델을 구축하고, 해당 모델을 Bootstrap resampling method를 통하여 모델 정교화를 진행하였다. Bootstrapping은 50번 진행하였으며, 해당 예측모형을 test set에 적용하여 모델해석 및 모델적합도를 평가하였다. Cox-proportional hazards model을 이용한 예측모델의 모델해석 및 모델적합도 평가는 Nomogram을 이용하였으며, 연구대상자 수가 너무 많아 Cox-proportional hazards model에서 calibration plot 및 Time dependent ROC curve를 통한 적합도 평가는 분석진행이 불가하였다. Logistic regression model의 모델적합도 평가는 Calibration plot, ROC curve을 통해 진행하였다. 데이터 전처리 및 T-test, ANOVA 분석은 SAS 9.4를 이용하였으며, 예측모델 구축 및 모델해석, 모델적합도 평가는 R 3.5.2를 이용하였다.

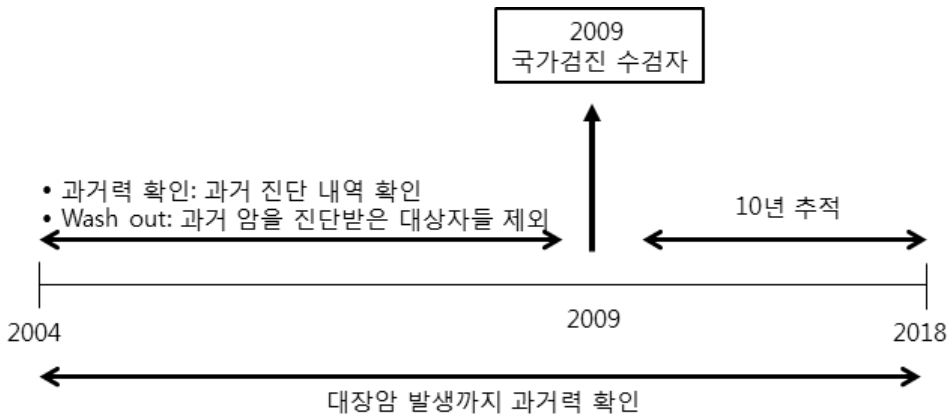
간암 발생 예측 모형은 Cox's proportional hazards model을 사용하였고, 다음과 같이 구하였다. 여기서  $\beta$ 는 각 위험요소의 회귀계수를,  $t$ 는 시간을 나타낸다.

$$h_i(t) = h_0(t) \exp(z), \quad z = \beta_1 Age_{1i} + \beta_2 Sex_{2i} + \beta_3 Screening\ result_{3i} + \dots + \beta_p X_{pi}$$

## 2. 대장암 질환 예측 연구

### (1) 일반검진자 대상 대장암 질환 예측을 위한 데이터셋 구축

2009년 국가 검진 수검자를 대상으로 2004년에서 2008년간 대장암 진단코드를 받은 대상자는 제외하고, 2018년까지 총 10년간의 간암 발생여부를 조사하였다. 대장암의 발생은 대장암 주상병코드(C18, C19, C20)를 기준으로 하였다. 결측치 및 검진 수치 이상치 상위1%, 하위 1%를 제외 한 후 총 8,787,004명을 대상으로 하였다.



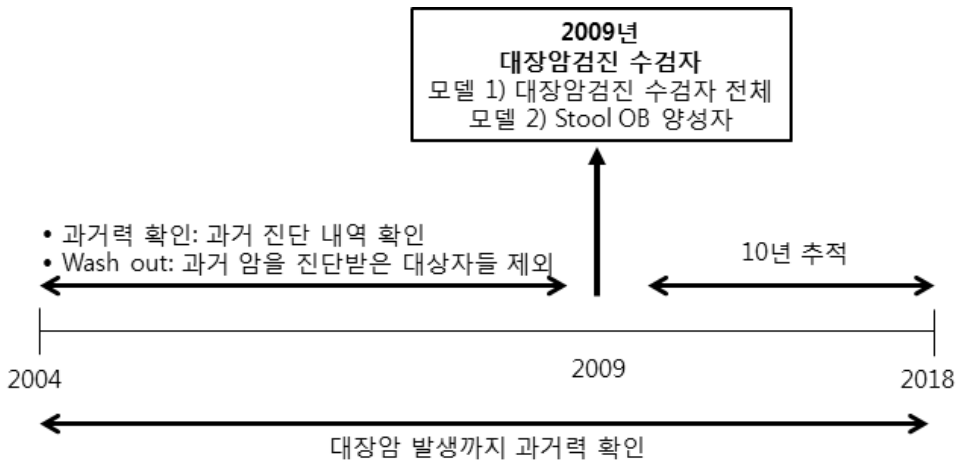
[그림 3-3] 일반검진자 대상 대장암 질환 예측 연구 대상자 선정

### (2) 대장암 검진자 대상 대장암 질환 예측을 위한 데이터셋 구축

2009년 대장암 검진 수검자를 대상으로 2004년에서 2008년간 대장암 진단코드를 받은 대상자는 제외하고, 2018년까지 총 10년간의 대장암 발생여부를 조사하였다. 대장암의 발생은 대장암 주상병코드(C18, C19, C20)를 기준으로 하였다. 결측치 및 검진수치 이상치 상위1%, 하위 1%를 제외 한 후 총 1,737,633명을 대상으로 하였다.

### (3) 분변잠혈검사 대상자 중 양성자의 대장암 질환 예측을 위한 데이터셋 구축

2009년 대장암 검진 수검자를 대상으로 2004년에서 2008년간 대장암 진단코드를 받은 대상자는 제외하고, 분변잠혈검사 양성자 중에서 2018년까지 총 10년간의 대장암 발생여부를 조사하였다. 결측치 및 검진수치 이상치 상위1%, 하위 1%를 제외 한 후 총 102,455명을 대상으로 하였다.



[그림 3-4] 대장암검진자 대상 대장암 질환 예측 연구 대상자 선정

#### (4) 대장암 예측 모델 개발에 포함될 변수 선정

이전 연구 사례에 근거한 대장암 발생의 위험 요소로는 연령, 성별, 흡연, 가족력, 음주, 비만 등의 요소가 공통적으로 지적되었으며, 자료가 있는 경우에는 이전의 대장내시경 시행 결과의 폴립 유무 등이 포함되어 있었다. 이 외에도 운동량, 식습관, 당뇨 등의 요소가 포함된 모델들이 있었다. 우리나라의 연구에서도 연령, 신장, BMI, 혈당, 콜레스테롤, 암의 가족력, 알코올 섭취, 육류 섭취 빈도 등이 포함되어 있었다.

##### ① 일반검진자 대상 위험인자 선정

간암 예측모델과 마찬가지로 과거 예측모형에서 사용된 위험요소들 중 국민건강보험 데이터에 존재하는 변수들을 본 연구에 포함시켰다. 최종적으로 이 연구에서는 나이, 성별, 수입, 체질량지수(kg/m<sup>2</sup>), 혈압(mmHg), 허리둘레(cm), AST(U/L), ALT(U/L), 콜레스테롤(mg/dL), 당뇨 유무, 고혈압 유무, 그리고 흡연, 음주, 운동 등이 위험요인으로 선정되었다. 나이는 40세 미만, 40-50, 50-60, 60-70, 70-80, 80세 이상으로 구분하였다, 소득은 의료급여층, 일반보험 대상자를 4분위하여 구분하였다. 체질량지수는 18.5 미만(저체중), 18.5~22.9, 23~24.9(과체중), 25~29.9(비만), 30이상(고도비만)으로 분류하였다. 당뇨병은 공복혈당 126 이상 또는 설문지에 의사에게 당뇨병 진단받은 경우로 정의하였다. 고혈압은 수축기 140이상 또는 이완기 90이상 또는 설문지상 고혈압을 진단받은 경우로 정의하였다. 그 외 대장암과 관련된 동반질환으로는 궤양성 대장염과 크론병을

기본적 특성에 포함시켰으나, 빈도수가 너무 적고 유의하지 않아 예측모델에는 포함하지 않았다.

생활습관은 다음과 같이 분류하였다. 흡연의 경우 비흡연, 과거흡연, 현재흡연 세 가지로 분류하였다. 음주의 경우 WHO 권고 기준에 따라 평균 하루 섭취 알코올 용량을 분류하였다. 남자의 경우  $0 < \text{저위험} < 40\text{g}$ ,  $40 \leq \text{중위험} < 60\text{g}$ ,  $60\text{g} \leq \text{고위험}$ 으로 분류하였다. 여자의 경우  $0 < \text{저위험} < 20\text{g}$ ,  $20 \leq \text{중위험} < 40\text{g}$ ,  $40\text{g} \leq \text{고위험}$ 으로 분류하였다. 평균 하루 섭취 알코올 용량은  $A \times 50\text{ml}(\text{소주한잔 용량}) \times 0.22(\text{소주 도수}) \times 0.8(\text{알콜 함량비}) \times B/7$ 으로 계산하였다. (A: 하루 섭취하는 잔수, B: 1주일에 마시는 횟수) 운동의 경우 WHO에서 제공하는 IPAQ-K 기준에 의해 분류하였다.  $0 \leq \text{저강도활동} < 599 \text{ MET} \times \text{min}$ ,  $600 \leq \text{중등도활동} < 2999 \text{ MET} \times \text{min}$ ,  $3000 \text{ MET} \times \text{min} < \text{고활동군}$ 으로 분류하였다. 운동은 설문지에 근거하여 격렬한 신체활동을 한 시간 이상 한 경우 주당 횟수  $\times 8.0 \text{ MET} \times 60$ 분으로 계산하고, 중등강도 신체활동을 한 시간 이상 한 경우 주당 횟수  $\times 4.0 \text{ MET} \times 60$ 분으로 계산, 저강도 활동을 30분 이상한 경우 걷기 주당 횟수  $\times 3.3 \text{ MET} \times 30$ 분으로 계산하였다.

## ② 대장암 검진자 대상 위험인자 선정

일반검진자에게 사용한 위험인자를 모두 포함시켰다. 추가로 국가에서 1년마다 시행하는 만 50세 이상 남녀 분변잠혈 검사 결과를 포함하였다. 폴립 과거력은 한번이라도 대장 내시경을 받았던 사람으로 혼란변수가 될 수 있으므로 분석에서 제외하였다.

## ③ 분변잠혈검사 양성 대상 위험인자 선정

일반검진자에게 사용한 위험인자를 모두 포함시켰다. 추가로 분변잠혈검사 양성 후 시행하는 내시경 여부와 이중조영검사 여부 결과를 포함하였다.

## (5) 분석 방법

암 발생에 따른 대상자의 기본적 특성 차이를 보기 위하여 Chi-square test 및 T-test 분석을 진행하였고, 암 경험이 없는 대상자에서 대장암이 발생하기 까지 소요 되는 시간과 발생 여부를 이용하여 건강검진정보와 대장암과의 연관성을 알아보기 위해 예측모델을 구축하였다. 전체 대상자를 70%의 Train set, 30%의 test set으로 나누어 분석하였다. Train set에서 Cox-proportional hazards model, Logistic regression 을 통하여 예측 모델을 구축하고, 해당 모델을 Bootstrap resampling



method를 통하여 모델 정교화를 진행하였다. Bootstrapping은 50번 진행하였으며, 해당 예측모델을 test set에 적용하여 모델해석 및 모델적합도를 평가하였다. Cox-proportional hazards model을 이용한 예측모델의 모델해석 및 모델적합도 평가는 Nomogram을 이용하였으며, 연구대상자 수가 너무 많아 Cox-proportional hazards model에서 calibration plot 및 Time dependent ROC curve를 통한 적합도 평가는 분석진행이 불가하였다. Logistic regression model의 모델적합도 평가는 Calibration plot, ROC curve을 통해 진행하였다. 데이터 전처리 및 T-test, ANOVA 분석은 SAS 9.4를 이용하였으며, 예측모델 구축 및 모델해석, 모델적합도 평가는 R 3.5.2를 이용하였다.

대장암 발생 예측 모형은 Cox's proportional hazards model을 사용하였고, 다음과 같이 구하였다. 여기서  $\beta$ 는 각 위험요소의 회귀계수를,  $t$ 는 시간을 나타낸다.

$$h_i(t) = h_0(t) \exp(z), \quad z = \beta_1 \text{Age}_{1i} + \beta_2 \text{Sex}_{2i} + \beta_3 \text{Screening result}_{3i} + \dots + \beta_p X_{pi}$$

### 3. 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과 확인

#### (1) 연구 대상자 및 위험요인 측정

연구 대상자는 모집기간에 연구에 동의한 국가검진 대상자 215명 중에서 사후관리 프로그램을 실제로 3개월간 사용한 63명을 대상으로 선정하였다. 건강검진의 신체 측정 자료를 통해 체질량지수, 허리둘레, 수축기혈압, 이완기혈압의 측정값을 이용하였으며, 건강검진의 혈액 검사 자료를 통해 공복혈당, 중성지방, 고밀도콜레스테롤의 결과치를 이용하였다. 또한 생활습관에 대해 운동 횟수, 흡연력, 음주력, 수면 시간, 스트레스 정도와 비만, 혈압, 고지혈증, 당뇨 관련 영양 습관에 대하여 확인하였다.

대사증후군 위험 요인으로 허리둘레, 수축기혈압, 이완기혈압, 공복혈당, 중성지방, 고밀도 콜레스테롤, 운동 횟수, 흡연력, 음주력, 수면시간, 스트레스 정도, 영양 습관을 선정하였다. 흡연력은 설문에서 비흡연자(non-smoker), 과거흡연자(ex-smoker)와 현재흡연자(current smoker)로 분류하여 정의하였다. 흡연력 유무의 경우, 비흡연자는 과거흡연자를 포함하였고 현재흡연자와 구분하였다. 음주는 음주 여부를 기준으로 분류하였고 음주자의 경우, 음주 횟수와 음주량(잔)을 비교하였다. 수면 시간은 7시간 이하와 8시간 이상을 합쳐 7-8시간과 구분하였고 스트레스 정도는 매우 적은 편(1), 적은 편(2), 보통(3), 많은 편(4), 매우 많은 편(5)으로 구분하여

1~5 숫자 척도를 사용하였다. 영양 습관은 비만, 혈압, 고지혈증, 당뇨 등 4가지 항목으로 나누어, 대사증후군으로 인해 유발되는 질환과 밀접한 관련이 있는 영양 습관으로 나누어 설문을 진행하였다.

(2) 대사증후군의 정의

대사증후군의 정의는 미국심장협회(AHA)와 국가심폐혈액연구소(NHLBI)에서 2005년에 종적 특성을 고려하여 발표한 modified ATP III criteria를 기준으로 하였다.

<표 3-2> modified ATP III criteria

Elevated waist circumference*	≥ 90 cm in men ≥ 85 cm in women
Elevated triglycerides	≥ 150 mg/dL (1.7 mmol/L) or On drug treatment for elevated triglycerides†
Reduced HDL-C	≤ 40 mg/dL (1.03 mmol/L) in men ≤ 50 mg/dL (1.3 mmol/L) in women
Elevated blood pressure	≥ 130 mm Hg systolic blood pressure / ≥ 85 mm Hg diastolic blood pressure or On antihypertensive drug treatment in a patient with a history
Elevated fasting glucose	≥ 100 mg/dL or On drug treatment for elevated glucose

<표 3-2>에 있는 modified ATP III criteria 위험요인 5개를 기준으로 0개 보유한 타입은 '정상군', 1-2개 보유한 타입은 '위험군', 3개 이상 보유한 타입은 '대사증후군'으로 분류하였다.

(3) 대사나이 타입 분류

대사나이에서 주민등록나이를 뺀 값을 기준으로 대사나이 타입을 나타내었다. -1 표준편차 값 이하인 경우 대사나이 '좋음' 타입, -1 표준편차 값에서 +1 표준편차 값 사이는 대사나이 '평균' 타입, +1 표준편차 값 이상인 경우 대사나이 '나쁨' 타입으로 분류하였다.

#### (4) 통계적 분석 방법

연구대상자 특성 파악은 사후관리 프로그램 사용 전과 후를 나누어 연속 변수는 평균과 표준 편차로 정리하였다. 또한, 사후관리 프로그램 사용 전과 후의 비교는 연속 변수의 경우 Paired T-Test, 명목 변수의 경우 Chi-square Test를 시행하였다. 이상의 분석은 파이썬(Python) 3.8.3 프로그램으로 수행하였으며, P값이 0.05 미만일 경우 통계적으로 의미 있는 결과로 판단하였다.

#### 4. 건강in 에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고

총 5회의 전문가 초청 포럼과 최종 토론회를 하였다. 포럼의 주제는 개인건강기록 경험과 전망, 데이터 관점의 전자의무기록, 진료정보교류 표준, 진료정보교류사업 현황과 방향, 외국사례를 통해 본 진료정보 활용 방향과 실행 전략, 국가건강정보포털의 현황과 방향, 비대면 환자교육의 내용이 다루어졌고, 미래의 건강in 의 방향으로 신뢰성 있는 대표적인 건강·의료 플랫폼의 구축과 함께 앱을 통한 개인 맞춤형서비스의 고도화를 위한 관련 토론을 하였다.

# 제4장

## 분석 결과

제1절 간암 발생 예측 연구	45
제2절 대장암 발생 예측 연구	67
제3절 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과 연구	88
제4절 건강in 에서의 개인별 맞춤형 통합 건강 관리의 고도화 및 방향성 제고	93



# 제4장 분석결과

## 제1절 간암 발생 예측 연구

### 1. 일반검진자 대상 간암 예측

#### (1) 일반검진 대상자의 특성

최대 10년의 추적관찰 기간 중에 전체 대상자 중 1.72%인 150,922명에서 간암이 발생하였다. 남성에서 더 발생하며, 나이는 70대에서 가장 많이 발생하였다. 경제수준은 의료 급여와 의료보험 상위 4분위에서 높게 나타났다. BMI, 허리둘레가 증가한 경우 높았다. 당뇨, 고혈압이 있는 경우 간암발생이 높았으며 간암 있는 군에서 혈압, 혈당 높게 측정되었다. 간수치 중 AST, ALT, rGT 모두 간암군에서 높았으며 지방간을 예측하는 fatty liver index 역시 간암군에서 높았다. 콜레스테롤 지표는 HDL 포함하여 LDL, Triglyceride, cholesterol 모두 낮게 나타났다. B형 간염 여부에서 가장 큰 빈도차이를 보였으며, C형 간염, 기타 간염, 간경변, 나이 등에서 큰 빈도차이를 보였다.

생활습관의 경우 흡연력은 경우 간암이 더 많이 발생하였다. 음주에서는 저위험 음주력이 있는 경우는 상대적으로 낮았으나 그 외에는 음주량과 비례하는 경향을 보였다. 운동량은 큰 상관성을 보이지 않았다.

<표 4-1> 일반검진 대상자의 간암 발생에 따른 기본적 특성

		전체	간암		p-value		
			미발생	발생			
		N	N	%	N	%	
나이	40 미만	2,728,698	2,707,750	99.23	20,948	0.77	<.0001
	40-49	2,359,714	2,324,281	98.5	35,433	1.5	
	50-59	1,881,530	1,838,153	97.69	43,377	2.31	
	60-69	1,188,273	1,154,670	97.17	33,603	2.83	
	70-79	551,588	535,823	97.14	15,765	2.86	
	80 이상	77,292	75,496	97.68	1,796	2.32	
	성별	Male	4,830,274	4,735,458	98.04	94,816	
Female		3,956,821	3,900,715	98.58	56,106	1.42	
소득	4분위	2,549,768	2,500,653	98.07	49,115	1.93	<.0001
	3분위	1,879,868	1,848,290	98.32	31,578	1.68	
	2분위	2,455,457	2,417,548	98.46	37,909	1.54	
	1분위	1,885,044	1,853,140	98.31	31,904	1.69	
	의료급여	16,958	16,542	97.55	416	2.45	
고혈압	정상	3,098,492	3,055,040	98.60	43,452	1.40	<.0001
	전기고혈압	3,700,270	3,639,509	98.36	60,761	1.64	
	고혈압	1,988,333	1,941,624	97.65	46,709	2.35	
당뇨병	정상	7,450,084	7,332,691	98.42	117,393	1.58	<.0001
	공복혈당장애	657,749	643,635	97.85	14,114	2.15	
	당뇨병	679,262	659,847	97.14	19,415	2.86	
Total cholesterol							
	평균 ± 표준편차	194.5 ± 33.39	194.6 ± 33.38		192.3 ± 33.93		<.0001
TG							
	평균 ± 표준편차	128.3 ± 75.18	128.3 ± 75.20		128.6 ± 73.75		0.0852
HDL							
	평균 ± 표준편차	54.87 ± 12.82	54.89 ± 12.82		53.99 ± 12.93		<.0001
LDL							
	평균 ± 표준편차	113.8 ± 30.79	113.8 ± 30.78		112.4 ± 31.31		<.0001
허리둘레							
	평균 ± 표준편차	63.71 ± 10.65	63.69 ± 10.65		64.63 ± 10.34		<.0001
고콜레스테롤	정상	6,096,721	5,994,081	98.32	102,640	1.68	<.0001
	고콜레스테롤	2,690,374	2,642,092	98.21	48,282	1.79	
BMI	낮음	230,953	228,342	98.87	2,611	1.13	<.0001
	정상	3,523,050	3,470,386	98.51	52,664	1.49	
	비만1단계	2,244,879	2,204,845	98.22	40,034	1.78	
	비만2단계	2,592,765	2,541,280	98.01	51,485	1.99	
흡연	비흡연	5,226,600	5,143,687	98.41	82,913	1.59	<.0001
	과거흡연	1,282,714	1,254,847	97.83	27,867	2.17	
	흡연	2,277,781	2,237,639	98.24	40,142	1.76	
음주	비음주군	4,553,900	4,472,619	98.22	81,281	1.78	<.0001
	저위험 음주군	3,730,152	3,671,840	98.44	58,312	1.56	
	중위험 음주군	290,975	285,578	98.15	5,397	1.85	
	고위험 음주군	212,068	206,136	97.2	5,932	2.80	
활동	고활동군	648,891	635,599	97.95	13,292	2.05	<.0001
	중등도 활동군	3,830,064	3,765,759	98.32	64,305	1.68	
	저활동군	4,308,140	4,234,815	98.3	73,325	1.70	

		전체	간암				p-value
			미발생		발생		
			N	N	%	N	
간경변	정상	7,640,063	7,528,784	98.54	111,279	1.46	<.0001
	진단	1,147,032	1,107,389	96.54	39,643	3.46	
B형 간염	정상	8,647,582	8,517,796	98.5	129,786	1.50	<.0001
	진단	139,513	118,377	84.85	21,136	15.15	
C형 간염	정상	8,755,667	8,608,356	98.32	147,311	1.68	<.0001
	진단	31,428	27,817	88.51	3,611	11.49	
기타 간염	정상	8,468,131	8,333,955	98.42	134,176	1.58	<.0001
	진단	318,964	302,218	94.75	16,746	5.25	
<b>총합</b>		<b>8,787,095</b>	<b>8,636,173</b>	<b>98.28</b>	<b>150,922</b>	<b>1.72</b>	

(\*The fatty liver index (FLI), which is an algorithm based on waist circumference, body mass index (BMI), triglyceride, and gamma-glutamyl-transferase (GGT), p value is less than 0.05 were determined to be statistically)

(2) cox-proportional hazards model 구축

예측모델에서는 빈도가 너무 작아서 불안정한 변수는 제외하였다. 예측모델에서 B형간염 진단 여부 및 나이에 따라 큰 격차를 보이는 것을 알 수 있으며, 고혈압과 운동량을 제외한 대부분의 변수들이 유의하였다.

<표 4-2> 일반검진 대상자의 간암 발생 예측모델 (생존분석)

		HR	95% CI		p-value
			Low	High	
나이	40 미만	1.000	-	-	
	40-49	1.940	1.921	1.958	<0.0001
	50-59	2.857	2.837	2.878	<0.0001
	60-69	3.575	3.557	3.592	<0.0001
	70-79	4.045	4.016	4.073	<0.0001
	80 이상	4.505	4.432	4.579	<0.0001
	성별	Male	1.000	-	-
	Female	0.694	0.672	0.716	<0.0001
소득	4분위	1.000	-	-	
	3분위	1.024	1.008	1.041	0.0113
	2분위	1.016	1.000	1.031	0.0756
	1분위	1.019	1.001	1.037	0.0463
	의료급여	1.162	1.040	1.283	0.0046
고혈압	정상	1.000	-	-	
	전기고혈압	0.989	0.977	1.002	0.1105
	고혈압	0.993	0.979	1.007	0.4435
당뇨병	정상	1.000	-	-	
	공복혈당장애	1.076	1.058	1.094	<0.0001
	당뇨병	1.198	1.178	1.218	<0.0001



	HR	95% CI		p-value
		Low	High	
허리둘레				
1cm당	1.004	1.003	1.005	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	0.900	0.887	0.913	<0.0001
흡연				
비흡연	1.000	-	-	
과거흡연	1.039	1.020	1.059	0.0001
흡연	1.121	1.104	1.138	<0.0001
음주				
비음주군	1.000	-	-	
저위험 음주군	0.944	0.928	0.960	<0.0001
중위험 음주군	1.079	1.043	1.115	<0.0001
고위험 음주군	1.364	1.337	1.391	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	0.998	0.977	1.020	0.822
저활동군	1.028	1.009	1.048	0.0074
간경변				
정상	1.000	-	-	
진단	1.564	1.550	1.578	<0.0001
B형 간염				
정상	1.000	-	-	
진단	7.895	7.879	7.911	<0.0001
C형 간염				
정상	1.000	-	-	
진단	2.533	2.483	2.583	<0.0001
기타 간염				
정상	1.000	-	-	
진단	1.862	1.840	1.884	<0.0001

(p value is less than 0.05, the results were determined to be statistically)

### (3) 일반검진 대상자의 예측모델 (Logistic regression)

생존분석과 비슷한 결과가 나타났다.

<표 4-3> 일반검진 대상자의 간암 발생 예측모델 (로지스틱 회귀분석)

	OR	95% CI		p-value
		Low	High	
나이				
40 미만	1.000	-	-	
40-49	1.944	1.920	1.967	<0.0001
50-59	2.891	2.874	2.909	<0.0001
60-69	3.593	3.570	3.617	<0.0001
70-79	3.825	3.795	3.856	<0.0001
80 이상	3.414	3.352	3.476	<0.0001
성별				
Male	1.000	-	-	
Female	0.714	0.692	0.736	<0.0001

	OR	95% CI		p-value
		Low	High	
소득				
4분위	1.000	-	-	
3분위	1.017	1.001	1.033	0.0224
2분위	1.007	0.990	1.025	0.311
1분위	1.008	0.988	1.028	0.3701
의료급여	1.164	1.055	1.274	0.0097
고혈압				
정상	1.000	-	-	
전기고혈압	0.991	0.977	1.004	0.1294
고혈압	0.996	0.979	1.012	0.4638
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.068	1.046	1.090	<0.0001
당뇨병	1.187	1.164	1.210	<0.0001
허리둘레				
1cm당	1.005	1.004	1.006	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	0.902	0.888	0.915	<0.0001
흡연				
비흡연	1.000	-	-	
과거흡연	1.037	1.015	1.059	0.0014
흡연	1.110	1.089	1.130	<0.0001
음주				
비음주군	1.000	-	-	
저위험음주군	0.956	0.942	0.970	<0.0001
중위험음주군	1.097	1.063	1.132	<0.0001
고위험음주군	1.404	1.365	1.443	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	0.996	0.972	1.020	0.6553
저활동군	1.020	0.999	1.041	0.08
간경변				
정상	1.000	-	-	
진단	1.571	1.555	1.588	<0.0001
B형 간염				
정상	1.000	-	-	
진단	8.781	8.764	8.799	<0.0001
C형 간염				
정상	1.000	-	-	
진단	3.037	2.977	3.097	<0.0001
기타 간염				
정상	1.000	-	-	
진단	1.921	1.900	1.942	<0.0001

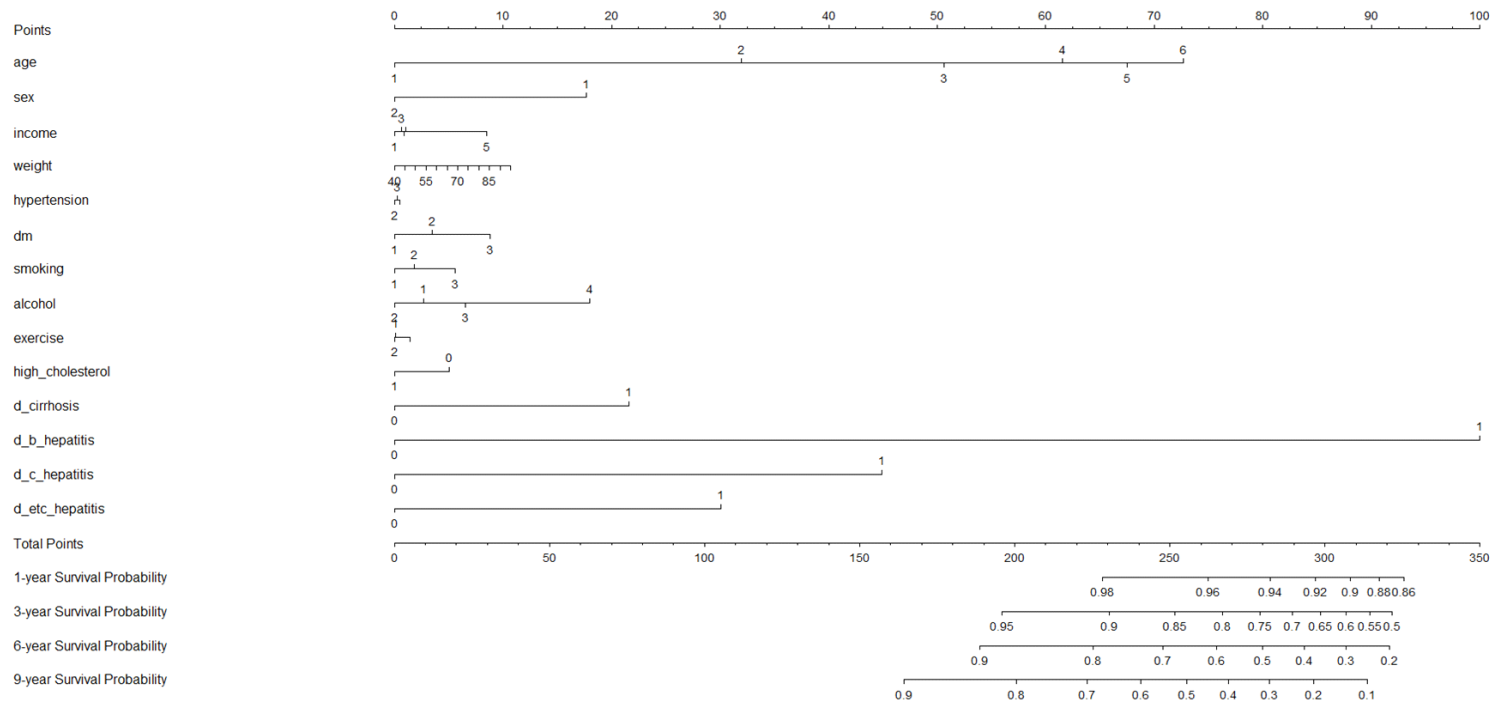
(Logistic regression was done. p value is less than 0.05, the results were determined to be statistically)

#### (4) 일반검진 대상자의 예측모델 노모그램(Nomogram)

노모그램은 질병과 위험요인 사이의 수치적 관계를 계산해 시각화하여 쉽고 직관적으로 이해할 수 있는 통계 도구로, 질병이나 사망에 관련된 위험요인을 선별하고 발생률을 예측하는데 이용된다. Point line(위험요인의 범주에 해당하는 점수 도출, 최상단에 배치), Risk factor line(사건 발생에 영향을 주는 위험요인의 범주의 점수를 나타내는 선으로 위험요인 개수만큼 위험요인 선을 구성), Probability line(사건의

발생 확률 도출), Total point line(통계적 모델을 바탕으로 계산하여 구성)으로 구성된다. 각각의 변수를 위로 선을 그으면 포인트를 알 수 있고 해당 포인트들을 전부 합친 것을 Total point에 대입하여 그 아래로 선을 그으면 1,3,6,9년 생존율을 알 수 있다.

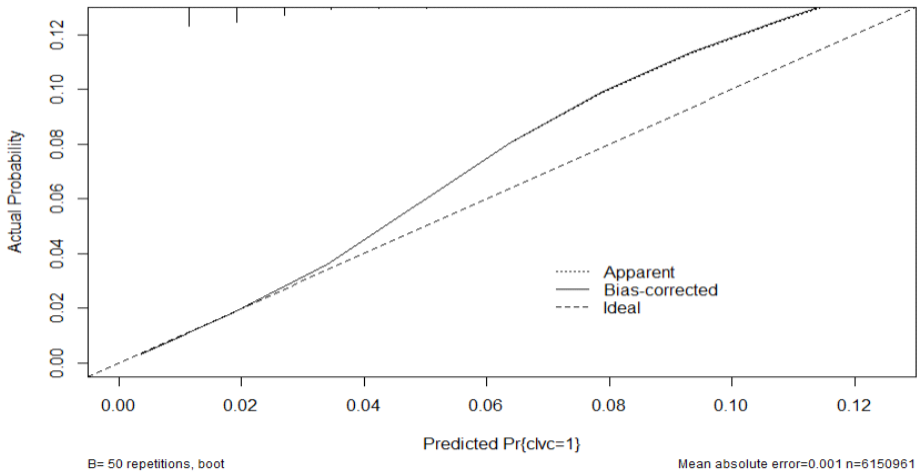
간경변, B형 간염, C형 간염, 기타 간염, 연령, 음주, 성 등에서 포인트가 상대적으로 높은 것을 알 수 있다.



[그림 4-1] 일반검진 대상자의 생존분석 nomogram

(5) 일반검진 대상자의 예측모델 Calibration plot

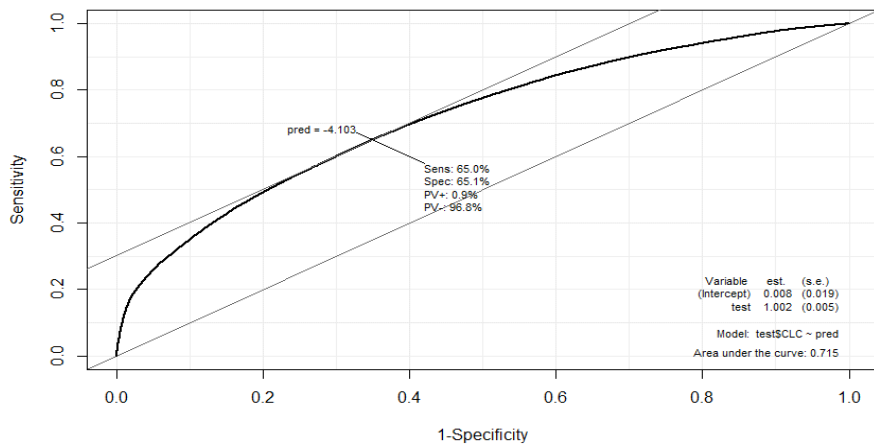
예측한 간암과 실제 관측된 간암이 얼마나 다른지 본 그래프이며 45도의 사선 점선에 가까울수록 예측과 실제 관측값이 같은 이상적인 그래프이다. 이 그래프에서는 확률이 높은 부분에서 예측확률보다 실제 관측값이 높아지는 것을 알 수 있다.



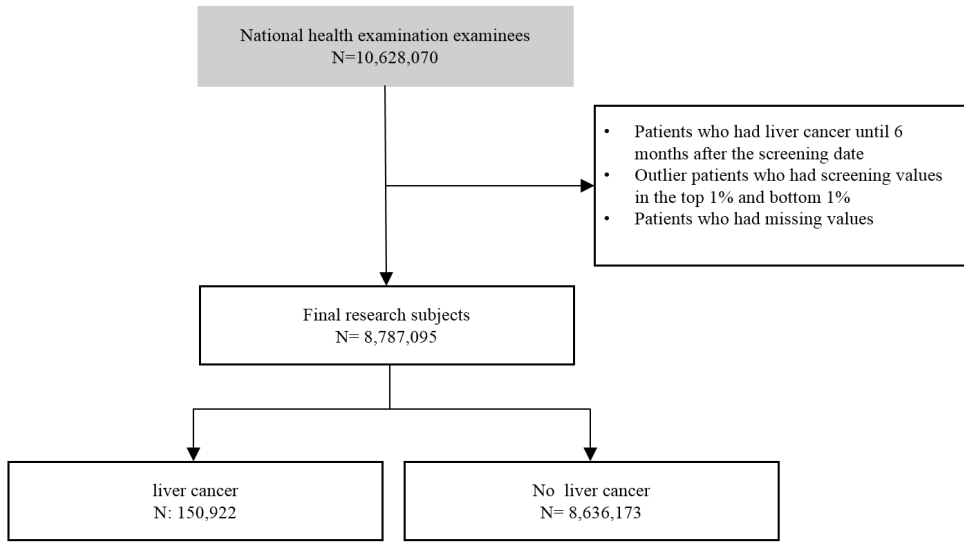
[그림 4-2] 일반검진 대상자의 calibration plot

(6) 일반검진 대상자의 예측모델 ROC curve

본 예측모델의 AUC값은 0.715로 분석되었다.



[그림 4-3] 일반검진 대상자의 ROC curve



[그림 4-4] 일반검진 대상자의 간암 발생 flow chart

## 2. 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측모델

(1) 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측 모델  
 나이가 가장 강력한 Hazard ratio를 나타냈다. 80세 이상은 HR이 5.361이었다. 흡연, 음주, 운동 등 생활습관이 악화될수록 HR를 증가하는 결과가 나타났다. 흡연하는 경우 HR는 1.157, 고위험 음주하는 경우 HR는 1.527으로 간암의 위험도가 올라갔다.

<표 4-4> 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측모델 (생존분석)

	HR	95% CI		p-value
		Low	High	
나이				
40 미만	1.000	-	-	
40-49	2.036	2.014	2.057	<0.0001
50-59	3.154	3.132	3.176	<0.0001
60-69	4.068	4.041	4.094	<0.0001
70-79	4.851	4.822	4.880	<0.0001
80 이상	5.291	5.222	5.361	<0.0001
성별				
Male	1.000	-	-	
Female	0.766	0.743	0.788	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.032	1.011	1.053	0.0016
2분위	1.029	1.011	1.047	0.0007
1분위	1.017	0.997	1.038	0.0404
의료급여	1.285	1.135	1.436	0.0013

	HR	95% CI		p-value
		Low	High	
고혈압				
정상	1.000	-	-	
전기고혈압	1.009	0.989	1.029	0.4829
고혈압	1.042	1.022	1.062	0.0002
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.113	1.085	1.142	<0.0001
당뇨병	1.242	1.217	1.267	<0.0001
허리둘레				
1cm당	1.005	1.005	1.006	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	0.928	0.913	0.943	<0.0001
흡연				
비흡연	1.000	-	-	
과거흡연	1.042	1.017	1.066	0.003
흡연	1.134	1.111	1.157	<0.0001
음주				
비음주군	1.000	-	-	
저위험 음주군	1.013	0.999	1.028	0.0387
중위험 음주군	1.189	1.149	1.228	<0.0001
고위험 음주군	1.487	1.448	1.527	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	0.951	0.920	0.983	0.0018
저활동군	0.981	0.950	1.013	0.2524

(p value is less than 0.05, the results were determined to be statistically)

(2) 간암검진 대상 기저질환 군 제외 후 간암 발생 예측 모델(로지스틱 회귀분석)  
생존분석과 비슷한 결과가 나타났다.

<표 4-5> 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 간암 발생 예측모델 (로지스틱 회귀분석)

	OR	95% CI		p-value
		Low	High	
나이				
40 미만	1.000	-	-	
40-49	2.033	2.006	2.059	<0.0001
50-59	3.156	3.129	3.182	<0.0001
60-69	4.032	4.002	4.062	<0.0001
70-79	4.528	4.490	4.566	<0.0001
80 이상	3.943	3.879	4.008	<0.0001
성별				
Male	1.000	-	-	
Female	0.787	0.764	0.809	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.030	1.007	1.053	0.0086
2분위	1.024	1.003	1.045	0.0156
1분위	1.009	0.985	1.033	0.365
의료급여	1.212	1.036	1.388	0.0186

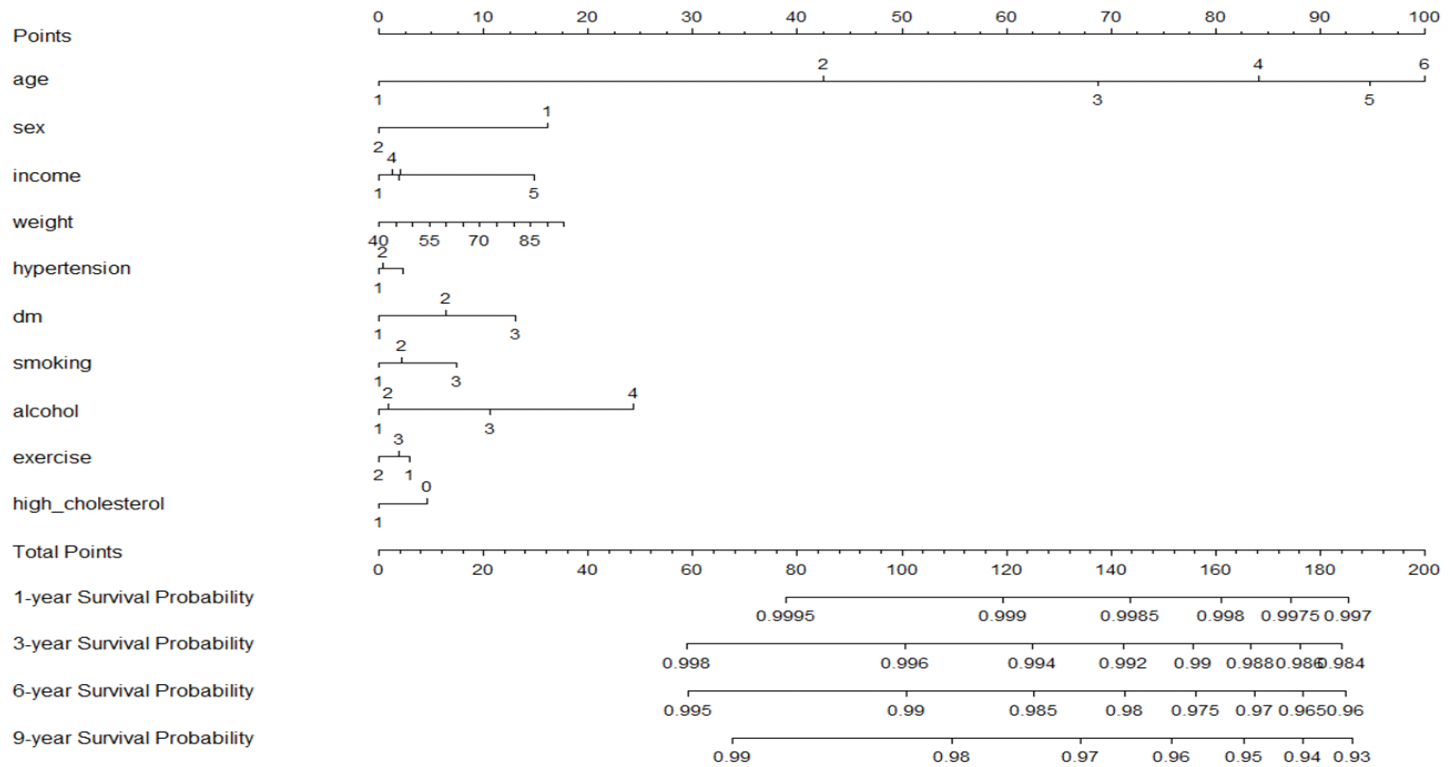
	OR	95% CI		p-value
		Low	High	
고혈압				
정상	1.000	-	-	
전기고혈압	1.009	0.991	1.028	0.4677
고혈압	1.039	1.012	1.066	0.0095
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.109	1.083	1.135	<0.0001
당뇨병	1.220	1.191	1.248	<0.0001
허리둘레				
1cm당	1.006	1.005	1.007	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	0.928	0.911	0.945	<0.0001
흡연				
비흡연	1.000	-	-	
과거흡연	1.038	1.015	1.060	0.0036
흡연	1.120	1.097	1.143	<0.0001
음주				
비음주군	1.000	-	-	
저위험 음주군	1.019	1.001	1.037	0.0216
중위험 음주군	1.197	1.155	1.239	<0.0001
고위험 음주군	1.495	1.450	1.540	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	0.946	0.916	0.975	0.0006
저활동군	0.970	0.940	1.000	0.0894

(p value is less than 0.05, the results were determined to be statistically)

### (3) 간암검진 대상 기저질환 군 제외 후 간암 발생 예측 모델 (nomogram)

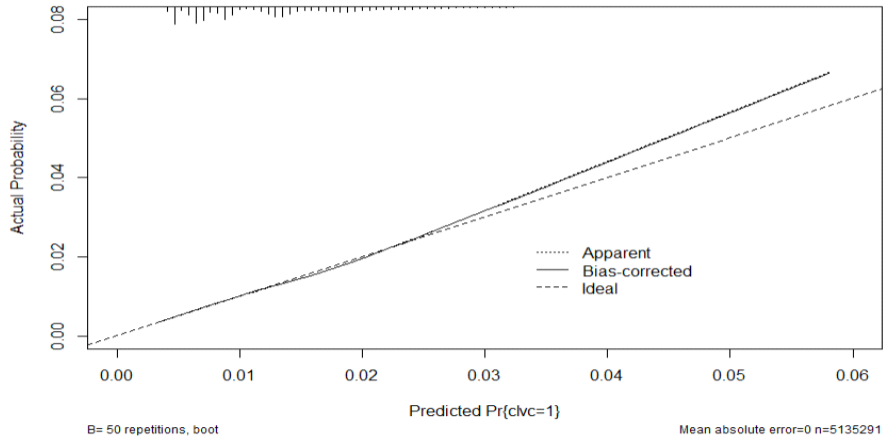
각각의 변수를 위로 선을 그으면 포인트를 알 수 있고 해당 포인트들을 전부 합친 것을 Total point에 대입하여 그 아래로 선을 그으면 1,3,6,9년 생존율을 알 수 있다. 연령, 음주, 성 등에서 포인트가 상대적으로 높은 것을 알 수 있다.





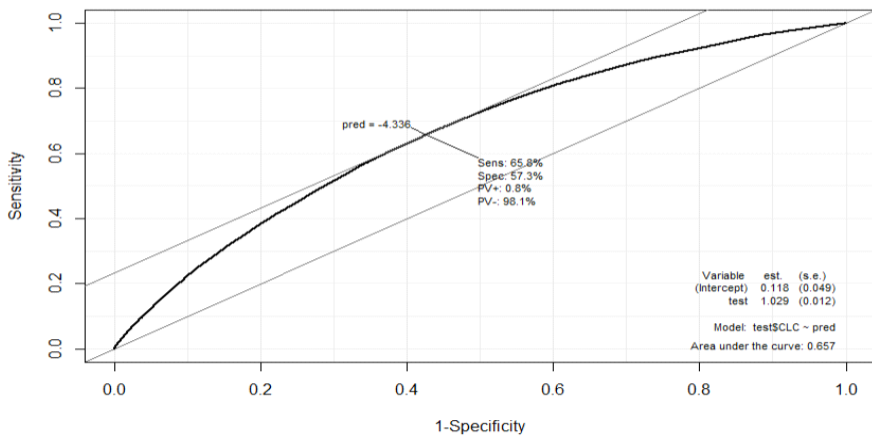
[그림 4-5] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 생존분석 nomogram

(4) 간암검진 대상 기저질환 군 제외 후 간암 발생 예측 모델 Calibration plot  
 예측한 간암과 실제 관측된 간암이 얼마나 다른지 본 그래프이며 45도의 사선 점선에 가까울수록 예측과 실제관측 값이 같은 이상적인 그래프이다. 이 그래프에서는 예측확률보다 실제 관측 값이 약간 높은 형태로 시작해서 약간 낮아졌다가 다시 높아지는 것을 알 수 있다.

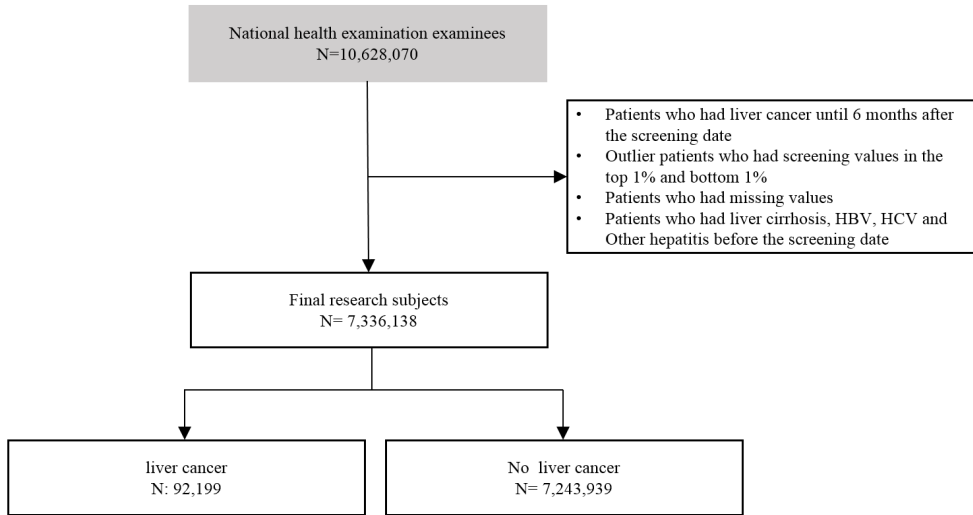


[그림 4-6] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 calibration plot

(5) 간암검진 대상 기저질환 군 제외 후 간암 발생 예측 모델 ROC curve  
 본 예측모델의 AUC 값은 0.657로 분석되었다.



[그림 4-7] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 ROC curve



[그림 4-8] 일반검진 대상자 중 간암검진 대상 기저질환 군 제외 후 flow chart

### 3. 간암검진자 대상 간암 예측

#### (1) 간암검진 대상자의 일반적 특성

최대 10년의 추적관찰 기간 중에 전체 대상자 중 10.14%인 17,390명이 간암 발생 간초음파검사 관찰소견에서 이형결절이 진단된 경우 가장 큰 빈도차이를 보였으나, HBV 42,077명 중 7,916명(18.81%), HCV 8,615명 중 1,637명(19%), 간경변 70,892명 중 8,863명(12.50%)에서 간암이 발생하였다. 나이는 60대에 가장 높은 발병을 보였다. 남성에서 간암 발생이 증가하였으며, 체중, 혈압 AST, ALT, rGT는 간암발생과 비례하여 증가하였다. 콜레스테롤 수치인 저밀도콜레스테롤, 중성지방, 콜레스테롤은 간암 군에서 더 낮게 나타났다. 고밀도 콜레스테롤 역시 낮은 경우 간암발생이 높았다. 생활습관에서는 고위험 음주 많이 상관성을 보였다. 이는 일반 검진자 특성과는 다른 경향을 보였다. <표 4-6>

<표 4-6> 간암검진 대상자의 간암 발생에 따른 기본적 특성

		전체		간암		p-value	
		N	N	%	N		%
나이	40-49	53,983	49,121	90.99	4,862	9.01	<.0001
	50-59	61,561	55,057	89.43	6,504	10.57	
	60-69	41,094	36,557	88.96	4,537	11.04	
	70-79	13,781	12,377	89.81	1,404	10.19	
	80 이상	1,060	977	92.17	83	7.83	
성별	Male	96,338	85,197	88.44	11,141	11.56	<.0001
	Female	75,141	68,892	91.68	6,249	8.32	
소득	4분위	21,623	19,448	89.94	2,175	10.06	0.8146
	3분위	39,146	35,130	89.74	4,016	10.26	
	2분위	44,172	39,686	89.84	4,486	10.16	
	1분위	32,661	29,396	90.00	3,265	10.00	
	의료급여	33,877	30,429	89.82	3,448	10.18	
고혈압	정상	48,859	43,797	89.64	5,062	10.36	0.0051
	전기고혈압	64,713	58,072	89.74	6,641	10.26	
	고혈압	57,907	52,220	90.18	5,687	9.82	
당뇨병	정상	131,230	118,087	89.98	13,143	10.02	<.0001
	공복혈당장애	16,646	14,981	90.00	1,665	10.00	
	당뇨병	23,603	21,021	89.06	2,582	10.94	
Total cholesterol	평균 ± 표준편차	193.6 ± 34.31	194.5 ± 34.27		184.6 ± 33.49		<.0001
TG	평균 ± 표준편차	130.8 ± 74.48	132.5 ± 75.31		114.8 ± 64.49		<.0001
HDL	평균 ± 표준편차	53.78 ± 13.05	53.8 ± 13.03		53.54 ± 13.20		0.0161
LDL	평균 ± 표준편차	113.4 ± 31.82	114 ± 31.91		107.8 ± 30.49		<.0001
허리둘레	평균 ± 표준편차	64.29 ± 9.94	64.21 ± 9.94		65.04 ± 9.8		<.0001
고콜레스테롤	정상	113,907	101,446	89.06	12,461	10.94	
	고콜레스테롤	57,572	52,643	91.44	4,929	8.56	
BMI	낮음	1,393	1,244	89.30	149	10.70	0.9285
	정상	55,898	50,265	89.92	5,633	10.08	
	비만전단계	47,394	42,565	89.81	4,829	10.19	
	비만1단계	62,195	55,886	89.86	6,309	10.14	
	비만2단계	4,599	4,129	89.78	470	10.22	
흡연	비흡연	36,066	32,043	88.85	4,023	11.15	<.0001
	과거흡연	32,328	28,621	88.53	3,707	11.47	
	흡연	103,085	93,425	90.63	9,660	9.37	
음주	비음주군	7,112	6,407	90.09	705	9.91	<.0001
	저위험 음주군	6,449	5,872	91.05	577	8.95	
	중위험 음주군	59,449	53,600	90.16	5,849	9.84	
	고위험 음주군	98,469	88,210	89.58	10,259	10.42	

	전체	간암				p-value
		미발생		발생		
	N	N	%	N	%	
운동						
고활동군	81,735	73,473	89.89	8,262	10.11	0.4389
중등도 활동군	72,693	65,261	89.78	7,432	10.22	
저활동군	17,051	15,355	90.05	1,696	9.95	
FLI정량						
평균 ± 표준편차	32.86 ± 23.84	32.71 ± 23.84		34.22 ± 23.76		<.0001
거친 예코상(간초음파검사 관찰소견)						
정상	156,949	142,915	91.06	14,034	8.94	<.0001
진단	14,530	11,174	76.9	3,356	23.10	
이형결절(간초음파검사 관찰소견)						
정상	170,969	153,759	89.93	17,210	10.07	<.0001
진단	510	330	64.71	180	35.29	
양성질환(간초음파검사 관찰소견)						
정상	102,887	90,818	88.27	12,069	11.73	<.0001
진단	68,592	63,271	92.24	5,321	7.76	
기타(간초음파검사 관찰소견)						
정상	149,114	134,200	90	14,914	10.00	<.0001
진단	22,365	19,889	88.93	2,476	11.07	
AFP						
음성	152,683	138,228	90.53	14,455	9.47	<.0001
양성	18,796	15,861	84.38	2,935	15.62	
간경변						
정상	100,587	92,060	91.52	8,527	8.48	<.0001
진단	70,892	62,029	87.5	8,863	12.50	
B형 간염						
정상	129,402	119,928	92.68	9,474	7.32	<.0001
진단	42,077	34,161	81.19	7,916	18.81	
C형 간염						
정상	162,864	147,111	90.33	15,753	9.67	<.0001
진단	8,615	6,978	81	1,637	19.00	
기타 간염						
정상	105,745	94,034	88.93	11,711	11.07	<.0001
진단	65,734	60,055	91.36	5,679	8.64	
총합	171,479	154,089	89.86	17,390	10.14	

(p value is less than 0.05 were determined to be statistically)

## (2) 간암검진자의 cox-proportional hazards model 구축

예측모델에서는 빈도가 너무 작아서 불안정한 변수는 제외하였다. 예측모델에서는 간초음파 관찰소견상 이형결절이 가장 큰 격차를 보이는 것으로 나타났으며, 일반 검진의 결과와 유사하게 B형 간염, C형 간염, 간경변 등의 기저질환 진단여부 및 나이에 따라 차이가 나타났다. 간암검진 대상자 예측모델에서 추가된 FLI, AFP, 간초음파 관찰소견(거친예코상, 이형결절) 등의 변수에서 큰 차이가 나타났다, 그 외에 대부분의 변수들이 유의하였다.

<표 4-7> 간암검진 대상자의 간암 발생 예측모델 (생존분석)

	HR	95% CI		p-value
		Low	High	
나이				
40-49	1.000	-	-	
50-59	1.243	1.200	1.287	<0.0001
60-69	1.361	1.301	1.421	<0.0001
70-79	1.411	1.337	1.485	<0.0001
80 이상	1.361	1.060	1.662	0.0379
성별				
Male	1.000	-	-	
Female	0.643	0.592	0.694	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.055	1.001	1.110	0.0694
2분위	1.013	0.956	1.071	0.1115
1분위	1.006	0.943	1.070	0.8083
의료급여	1.061	0.982	1.139	0.096
고혈압				
정상	1.000	-	-	
전기고혈압	0.939	0.890	0.988	0.0094
고혈압	0.886	0.830	0.943	<0.0001
당뇨병				
정상	1.000	-	-	
공복혈당장애	0.966	0.895	1.037	0.3687
당뇨병	1.081	1.033	1.129	0.0012
허리둘레				
1cm당	0.992	0.990	0.995	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	0.746	0.702	0.789	<0.0001
흡연				
비흡연	1.000	-	-	
과거흡연	0.960	0.900	1.020	0.228
흡연	0.950	0.891	1.010	0.0899
음주				
비음주군	1.000	-	-	
저위험 음주군	0.901	0.795	1.007	0.091
중위험 음주군	1.030	0.944	1.116	0.4656
고위험 음주군	1.229	1.141	1.317	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	1.020	0.988	1.051	0.2899
저활동군	0.910	0.854	0.967	0.001
FLI				
1 증가시	1.009	1.008	1.010	<0.0001
거친 예코상(간초음파검사 관찰조건)				
정상	1.000	-	-	
진단	1.983	1.922	2.044	<0.0001

	HR	95% CI		p-value
		Low	High	
이형결절(간초음파검사 관찰조건)				
정상	1.000	-	-	
진단	3.072	2.887	3.257	<0.0001
양성질환(간초음파검사 관찰조건)				
정상	1.000	-	-	
진단	0.803	0.759	0.847	<0.0001
기타(간초음파검사 관찰조건)				
정상	1.000	-	-	
진단	1.150	1.089	1.211	<0.0001
AFP				
음성	1.000	-	-	
양성	1.568	1.512	1.623	<0.0001
간경변				
정상	1.000	-	-	
진단	1.400	1.358	1.443	<0.0001
B형 간염				
정상	1.000	-	-	
진단	2.648	2.610	2.686	<0.0001
C형 간염				
정상	1.000	-	-	
진단	2.116	2.042	2.189	<0.0001
기타 간염				
정상	1.000	-	-	
진단	0.924	0.886	0.961	0.0001

(p value is less than 0.05, the results were determined to be statistically)

### (3) 간암검진 대상자의 예측모델 (Logistic regression)

생존분석과 비슷한 결과가 나타났다.

<표 4-8> 간암검진 대상자의 간암 발생 예측모델 (로지스틱 회귀분석)

	OR	95% CI		p-value
		Low	High	
나이				
40-49	1.000	-	-	
50-59	1.256	1.198	1.313	<0.0001
60-69	1.363	1.308	1.418	<0.0001
70-79	1.303	1.219	1.387	<0.0001
80 이상	1.063	0.773	1.353	0.5393
성별				
Male	1.000	-	-	
Female	0.651	0.595	0.707	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.067	0.995	1.139	0.1096
2분위	1.036	0.961	1.111	0.5234
1분위	1.030	0.943	1.117	0.6159
의료급여	1.077	1.001	1.153	0.0789

	OR	95% CI		p-value
		Low	High	
고혈압				
정상	1.000	-	-	
전기고혈압	0.935	0.893	0.976	0.0012
고혈압	0.879	0.828	0.930	<0.0001
당뇨병				
정상	1.000	-	-	
공복혈당장애	0.961	0.889	1.033	0.2642
당뇨병	1.072	1.014	1.130	0.0339
허리둘레				
1cm당	0.994	0.991	0.997	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	0.735	0.683	0.787	<0.0001
흡연				
비흡연	1.000	-	-	
과거흡연	0.978	0.916	1.040	0.486
흡연	0.962	0.908	1.016	0.1774
음주				
비음주군	1.000	-	-	
저위험 음주군	0.933	0.819	1.048	0.201
중위험 음주군	1.042	0.936	1.148	0.4149
고위험 음주군	1.239	1.137	1.341	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	1.019	0.979	1.060	0.3614
저활동군	0.907	0.833	0.981	0.0089
FLI				
1 증가시	1.009	1.008	1.010	<0.0001
거친 에코상(간초음파검사 관찰소견)				
정상	1.000	-	-	
진단	2.194	2.136	2.251	<0.0001
이형결절(간초음파검사 관찰소견)				
정상	1.000	-	-	
진단	3.448	3.233	3.663	<0.0001
양성질환(간초음파검사 관찰소견)				
정상	1.000	-	-	
진단	0.808	0.758	0.857	<0.0001
기타(간초음파검사 관찰소견)				
정상	1.000	-	-	
진단	1.173	1.110	1.235	<0.0001
AFP				
음성	1.000	-	-	<0.0001
양성	1.601	1.550	1.653	
간경변				
정상	1.000	-	-	
진단	1.421	1.377	1.464	<0.0001
B형 간염				
정상	1.000	-	-	
진단	2.929	2.887	2.971	<0.0001
C형 간염				
정상	1.000	-	-	
진단	2.356	2.285	2.427	<0.0001
기타 간염				
정상	1.000	-	-	
진단	0.929	0.894	0.964	<0.0001

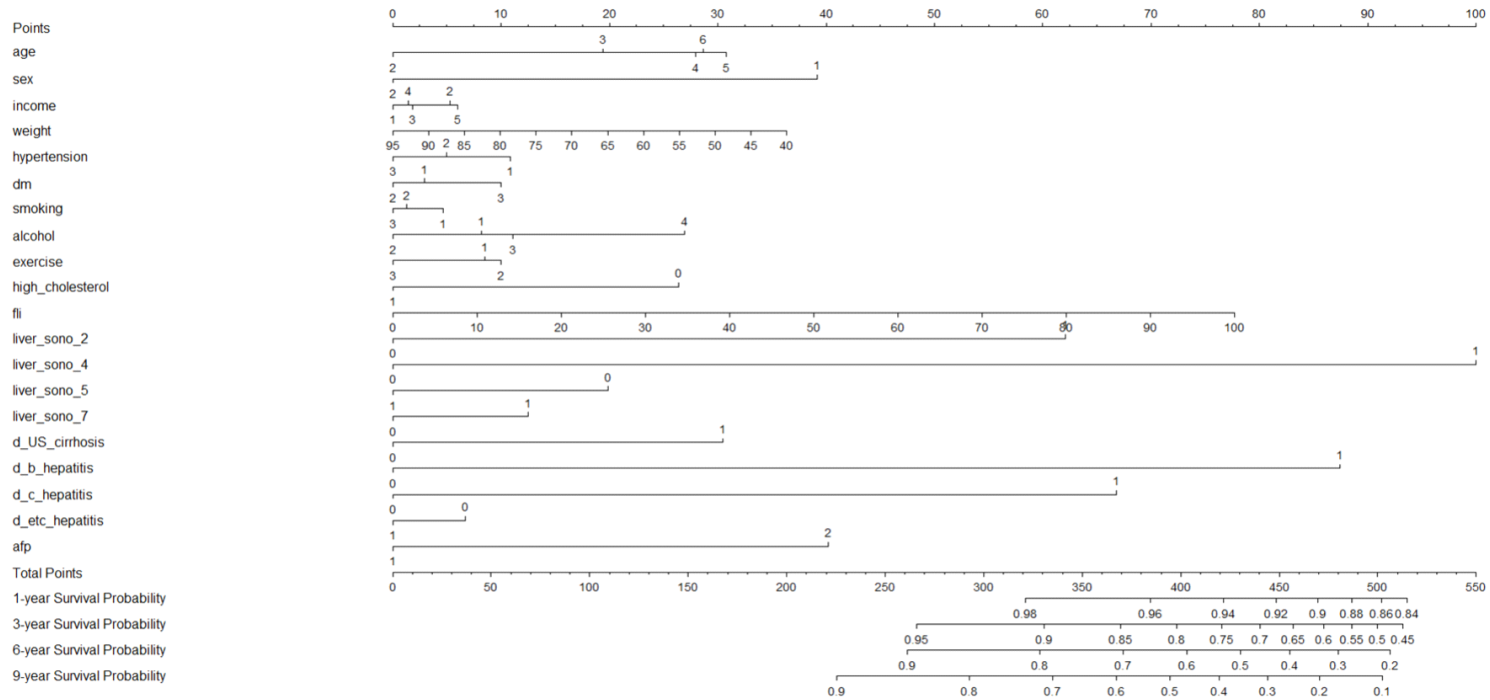
(p value is less than 0.05, the results were determined to be statistically)



#### (4) 간암검진 대상자의 예측모델 Nomogram

각각의 변수를 위로 선을 그으면 포인트를 알 수 있고 해당 포인트들을 전부 합친 것을 Total point에 대입하여 그 아래로 선을 그으면 1,3,6,9년 생존율을 알 수 있다.

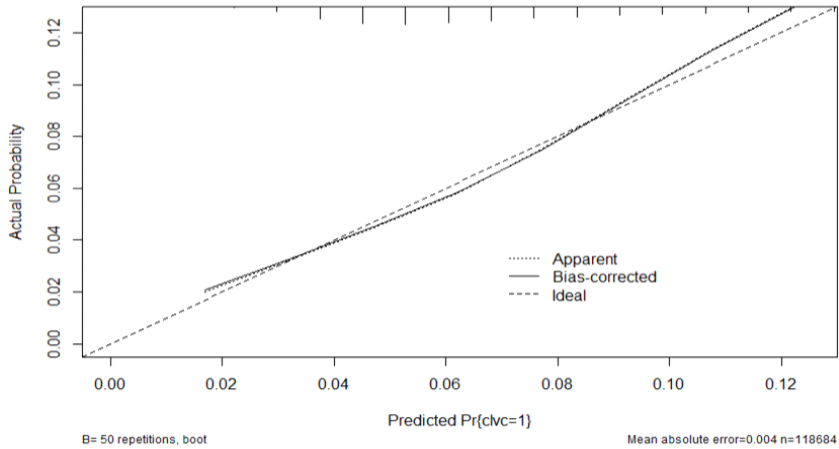
간초음파 관찰소견에서 이형결절로 진단받은 경우 점수가 가장 높은 것을 확인 할 수 있다. 그 외 B형 간염, C형 간염, 거친에코상(간초음파 관찰소견), AFP 등에서 점수가 상대적으로 높은 것을 확인 할 수 있다. 간암 검진 관련 변수가 추가 되며, 일반검진에서 점수가 높았던 연령과 기타 간염의 점수가 상대적으로 낮아진 것을 확인 할 수 있고, 반대로 성, 허리둘레, high\_cholesterol 등은 점수가 상대적으로 높아진 것을 확인 할 수 있다.



[그림 4-9] 간암검진 대상자의 생존분석 nomogram

(5) 간암검진 대상자의 예측모델 Calibration plot

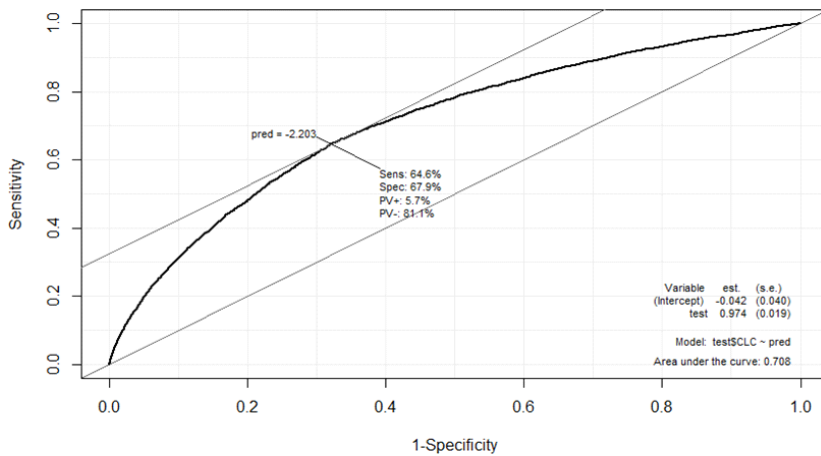
예측한 간암과 실제 관측된 간암이 얼마나 다른지 본 그래프이며 45도의 사선 점선에 가까울수록 예측과 실제 관측값이 같은 이상적인 그래프이다. 이 그래프에서는 예측확률보다 실제 관측값이 약간 높은 형태로 시작해서 약간 낮아졌다가 다시 높아지는 것을 알 수 있다.



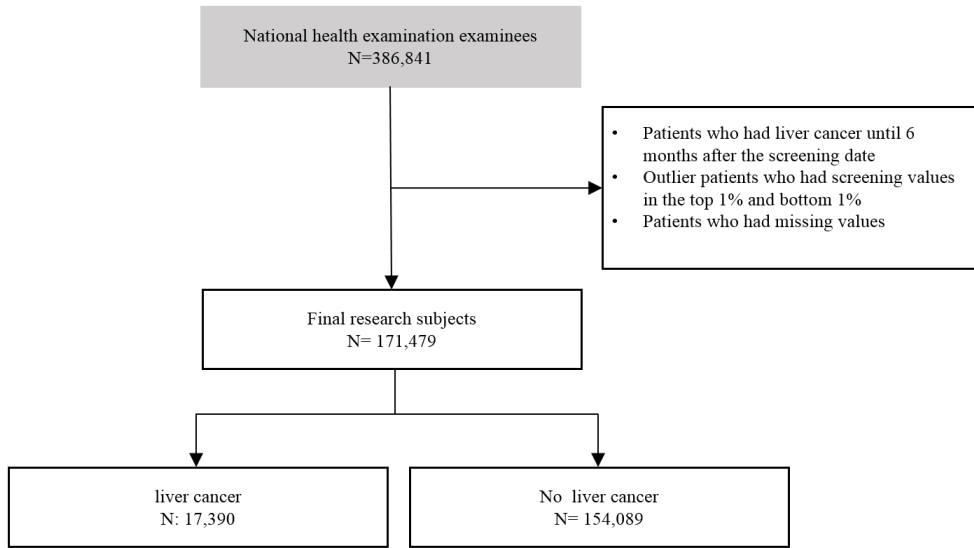
[그림 4-10] 간암검진 대상자의 calibration plot

(6) 간암검진 대상자의 예측모델 ROC curve

본 예측모델의 AUC값은 0.708로 분석되었다.



[그림 4-11] 간암검진 대상자의 ROC curve



[그림 4-12] 간암 검진 대상자의 간암 발생 flow chart

## 제2절 대장암 발생 예측 연구

### 1. 일반검진자 대상 대장암 예측

#### (1) 일반검진 대상자의 특성

최대 10년의 추적관찰 기간 중에 전체 대상자 중 1.2%인 105,454명에서 대장암이 발생하였다. 70대에서 발생률이 가장 높았고 여성보다 남성에서 더 발생하였다. 경제 수준은 의료 급여에서 발생률이 가장 높게 나타났다. 고혈압, 당뇨병이 있는 경우 대장암 발생률이 높았다. 콜레스테롤 지표는 Total cholesterol, TG, LDL 모두 대장암 발생 군에서 높게 측정되었고, HDL은 대장암 발생군에서 낮게 측정되었으며, 고콜레스테롤인 경우 대장암 발생률이 높았다. 비만 1단계와 2단계에서 높은 발생률을 보였다.

생활습관의 경우 과거 흡연 경험이 있는 군에서 대장암이 더 많이 발생하였다. 음주에서는 고위험 음주력이 있는 경우 대장암 발생률이 높았으며, 운동량은 고활동 군에서 대장암 발생률이 높았다. 그 외에 궤양성 대장염과 크론병을 진단 받은 경우 대장암 발생률이 높았다.

<표 4-9> 일반검진 대상자의 대장암 발생에 따른 기본적 특성

	전체	대장암				p-value
		미발생		발생		
	N	N	%	N	%	
나이						
40 미만	2,724,284	2,717,389	99.75	6,895	0.25	<.0001
40-49	2,365,630	2,347,593	99.24	18,037	0.76	
50-59	1,887,775	1,858,864	98.47	28,911	1.53	
60-69	1,187,224	1,155,973	97.37	31,251	2.63	
70-79	547,538	529,429	96.69	18,109	3.31	
80 이상	76,553	74,302	97.06	2,251	2.94	
성별						
Male	4,833,900	4,769,408	98.67	64,492	1.33	<.0001
Female	3,955,104	3,914,142	98.96	40,962	1.04	
소득						
의료급여	17,029	16,776	98.51	253	1.49	<.0001
1분위	1,885,636	1,861,677	98.73	23,959	1.27	
2분위	2,456,694	2,431,910	98.99	24,784	1.01	
3분위	1,880,719	1,859,613	98.88	21,106	1.12	
4분위	2,548,926	2,513,574	98.61	35,352	1.39	
고혈압						
정상	3,099,560	3,073,763	99.17	25,797	0.83	<.0001
전기고혈압	3,701,679	3,661,445	98.91	40,234	1.09	
고혈압	1,987,765	1,948,342	98.02	39,423	1.98	
당뇨병						
정상	7,450,816	7,371,837	98.94	78,979	1.06	<.0001
공복혈당 장애	658,366	647,573	98.36	10,793	1.64	
당뇨병	679,822	664,140	97.69	15,682	2.31	
Total cholesterol						
평균 ± 표준편차	194.56 ± 33.35	194.5 ± 33.33		197.4 ± 33.94		<.0001
TG						
평균 ± 표준편차	128.29 ± 75.17	128.2 ± 75.14		138.4 ± 76.36		<.0001
HDL						
평균 ± 표준편차	54.88 ± 12.82	54.89 ± 12.82		53.52 ± 12.71		<.0001
LDL						
평균 ± 표준편차	113.84 ± 30.76	113.8 ± 30.74		115.9 ± 31.69		<.0001
허리둘레						
평균 ± 표준편차	63.72 ± 10.65	63.72 ± 10.65		63.73 ± 10.02		<.0001
고콜레스테롤						
정상	6,099,270	6,031,968	98.9	67,302	1.10	<.0001
고콜레스테롤	2,689,734	2,651,582	98.58	38,152	1.42	
BMI						
낮음	230,377	228,693	99.27	1,684	0.73	<.0001
정상	3,520,723	3,484,418	98.97	36,305	1.03	
비만1단계	2,245,999	2,217,220	98.72	28,779	1.28	
비만2단계	2,596,057	2,560,087	98.61	35,970	1.39	
비만3단계	195,848	193,132	98.61	2,716	1.39	
흡연						
비흡연	5,226,677	5,166,763	98.85	59,914	1.15	<.0001
과거흡연	1,282,567	1,262,007	98.4	20,560	1.60	
흡연	2,279,760	2,254,780	98.9	24,980	1.10	
음주						
비음주군	4,553,898	4,495,496	98.72	58,402	1.28	<.0001
저위험 음주군	3,730,882	3,691,002	98.93	39,880	1.07	
중위험 음주군	291,401	287,823	98.77	3,578	1.23	
고위험 음주군	212,823	209,229	98.31	3,594	1.69	

	전체	대장암				p-value
		미발생		발생		
		N	%	N	%	
운동						
고활동군	711,841	701,097	98.49	10,744	1.51	<.0001
중등도 활동군	3,332,146	3,294,159	98.86	37,987	1.14	
저활동군	4,745,017	4,688,294	98.8	56,723	1.20	
궤양성 대장염						
정상	8,777,459	8,672,204	98.8	105,255	1.20	<.0001
진단	11,545	11,346	98.28	199	1.72	
크론병						
정상	8,786,920	8,681,511	98.8	105,409	1.20	<.0001
진단	2,084	2,039	97.84	45	2.16	
총합	8,789,004	8,683,550	98.80	105,454	1.20	

(2) cox-proportional hazards model 구축

예측모델에서는 빈도가 너무 작아서 불안정한 변수는 제외하였다. 예측모델에서 나이에 따라 가장 큰 격차를 보이는 것을 알 수 있으며, 그 외 대부분의 변수들이 유의하였고 그 중 당뇨병, 음주에서 가장 큰 차이를 보였다.

<표 4-10> 일반검진 대상자의 대장암 발생 예측모델 (생존분석)

	HR	95% CI		p-value
		Low	High	
나이				
40 미만	1.000	-	-	
40-49	3.173	3.126	3.220	<0.0001
50-59	6.539	6.495	6.583	<0.0001
60-69	11.331	11.288	11.373	<0.0001
70-79	15.427	15.376	15.477	<0.0001
80 이상	17.051	16.950	17.152	<0.0001
성별				
Male	1.000	-	-	
Female	0.717	0.683	0.751	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.041	1.003	1.079	0.0527
2분위	1.001	0.977	1.026	0.9115
1분위	1.054	1.028	1.081	<0.0001
의료급여	1.246	1.052	1.439	0.0228
고혈압				
정상	1.000	-	-	
전기고혈압	1.025	1.002	1.048	0.0363
고혈압	1.068	1.037	1.098	<0.0001
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.113	1.074	1.152	<0.0001
당뇨병	1.233	1.198	1.269	<0.0001

	HR	95% CI		p-value
		Low	High	
허리둘레				
1cm당	1.003	1.002	1.005	0.0002
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	1.026	1.001	1.050	0.049
흡연				
비흡연	1.000	-	-	
과거흡연	1.132	1.098	1.165	<0.0001
흡연	1.111	1.076	1.146	<0.0001
음주				
비음주군	1.000	-	-	
저위험 음주군	1.072	1.043	1.101	<0.0001
중위험 음주군	1.234	1.175	1.294	<0.0001
고위험 음주군	1.263	1.205	1.321	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	1.017	0.983	1.052	0.3376
저활동군	1.024	0.989	1.060	0.1979

(3) 일반검진 대상자의 예측모델 (Logistic regression)

생존분석과 비슷한 결과가 나타났다.

<표 4-11> 일반검진 대상자의 대장암 발생 예측모델 (로지스틱 회귀분석)

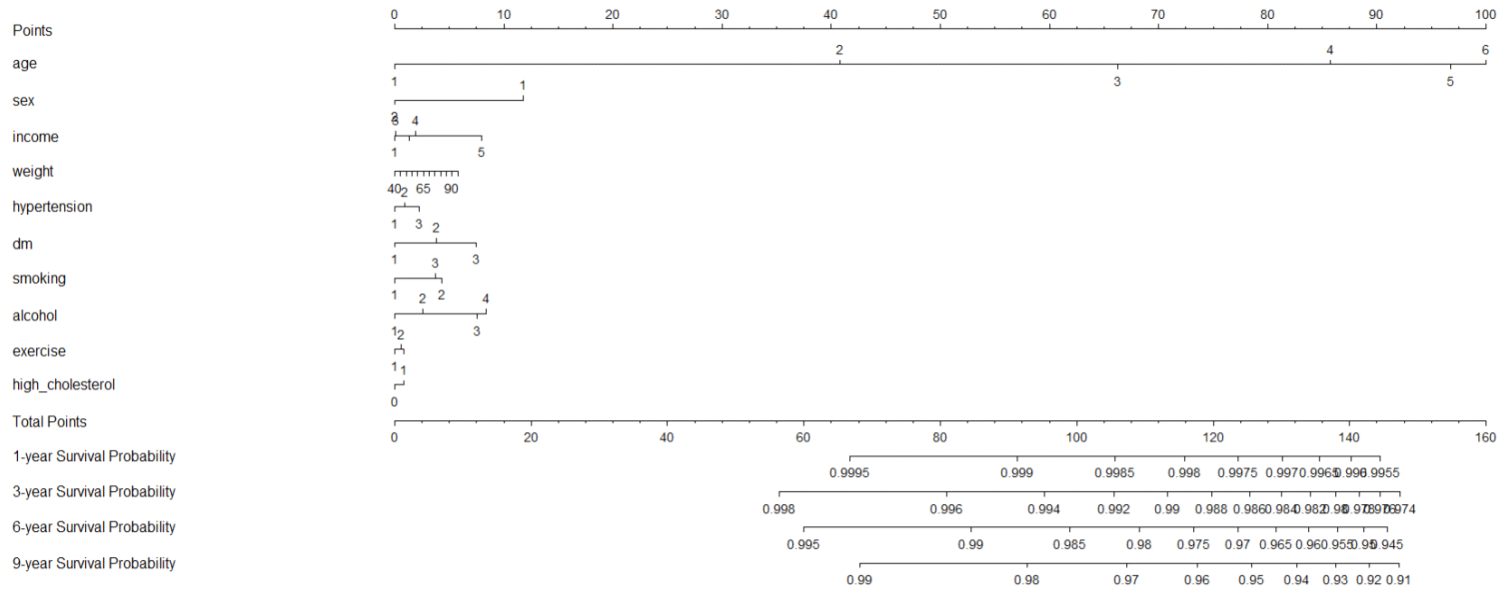
	OR	95% CI		p-value
		Low	High	
나이				
40 미만	1.000	-	-	
40-49	3.176	3.128	3.224	<0.0001
50-59	6.556	6.503	6.608	<0.0001
60-69	11.268	11.220	11.316	<0.0001
70-79	14.536	14.479	14.594	<0.0001
80 이상	12.821	12.714	12.928	<0.0001
성별				
Male	1.000	-	-	
Female	0.747	0.717	0.777	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.033	1.001	1.064	0.0269
2분위	0.994	0.967	1.021	0.791
1분위	1.048	1.019	1.077	0.0017
의료급여	1.204	0.974	1.435	0.1294
고혈압				
정상	1.000	-	-	
전기고혈압	1.026	0.996	1.055	0.0869
고혈압	1.062	1.031	1.094	0.0001
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.105	1.070	1.140	<0.0001
당뇨병	1.208	1.180	1.236	<0.0001

	OR	95% CI		p-value
		Low	High	
허리둘레				
1cm당	1.004	1.003	1.005	<0.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	1.029	1.005	1.052	0.031
흡연				
비흡연	1.000	-	-	
과거흡연	1.125	1.099	1.150	<0.0001
흡연	1.092	1.062	1.122	<0.0001
음주				
비음주군	1.000	-	-	
저위험 음주군	1.094	1.070	1.119	<0.0001
중위험 음주군	1.249	1.183	1.316	<0.0001
고위험 음주군	1.267	1.206	1.328	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	1.014	0.973	1.054	0.4281
저활동군	1.014	0.979	1.048	0.3788

(4) 일반검진 대상자의 예측모델 노모그램(Nomogram)

각각의 변수를 위로 선을 그으면 포인트를 알 수 있고 해당 포인트들을 전부 합친 것을 Total point에 대입하여 그 아래로 선을 그으면 1,3,6,9년 생존율을 알 수 있다. 나이, 성별, 소득, 음주, 당뇨병 등에서 포인트가 상대적으로 높은 것을 알 수 있다.

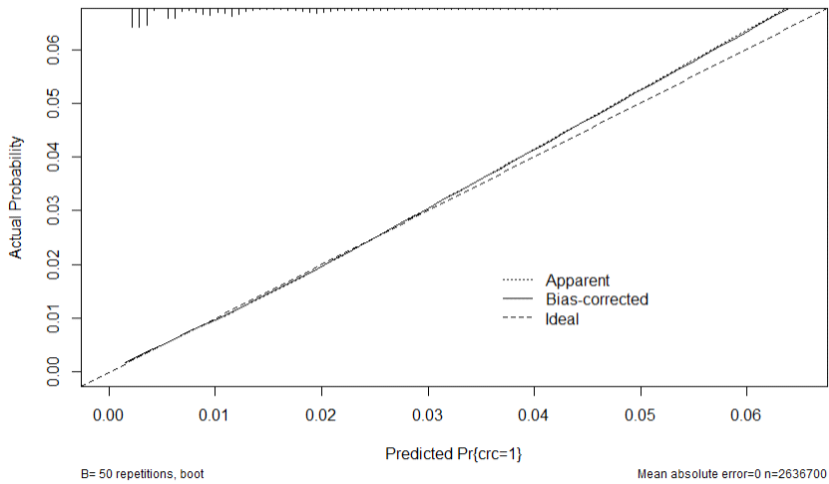




[그림 4-13] 일반검진 대상자의 대장암 발생 예측모델 Nomogram

(5) 대상자의 예측모델 Calibration plot

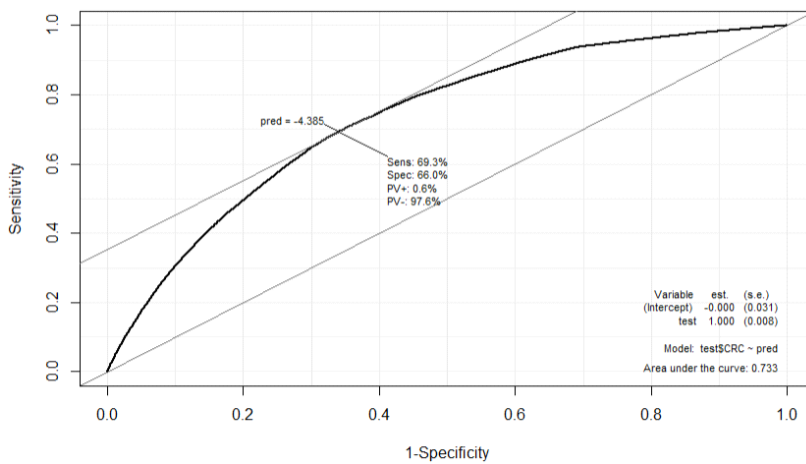
예측한 대장암과 실제 관측된 대장암이 얼마나 다른지 본 그래프이며 45도의 사선 점선에 가까울수록 예측과 실제 관측 값이 같은 이상적인 그래프이다. 이 그래프에서는 확률이 높은 부분에서 예측확률보다 실제 관측 값이 높아지는 것을 알 수 있다.



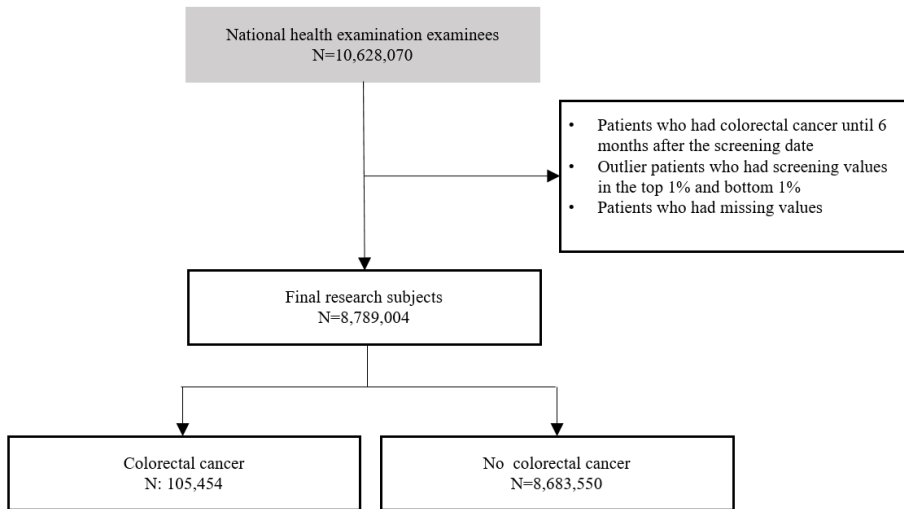
[그림 4-14] 일반검진 대상자의 대장암 발생 예측모델 Calibration plot

(6) 일반검진 대상자의 예측모델 ROC curve

본 예측모델의 AUC 값은 0.733으로 분석되었다.



[그림 4-15] 일반검진 대상자의 대장암 발생 예측모델 ROC curve



[그림 4-16] 일반검진 대상자의 대장암 발생 flow chart

## 2. 대장암 검진자 대상 대장암 발생 예측

### (1) 대장암검진 대상자의 일반적 특성

최대 10년의 추적관찰 기간 중에 전체 대상자 중 2.21%인 38,358명에서 대장암이 발생하였다. 80대에서 발생률이 가장 높았고 여성보다 남성에서 더 발생하였다. 경제 수준은 의료 급여에서 발생률이 가장 높게 나타났다. 고혈압, 당뇨병이 있는 경우 대장암 발생률이 높았다. 콜레스테롤 지표는 Total cholesterol, HDL, LDL 모두 대장암 발생 군에서 낮게 측정되었고, TG는 대장암 발생군에서 높게 측정되었으며, 고콜레스테롤인 경우 대장암 발생률이 높았다. 비만 1단계에서 가장 높은 발생률을 보였다.

생활습관의 경우 과거 흡연 경험이 있는 군에서 대장암이 더 많이 발생하였다. 음주에서는 고위험 음주력이 있는 경우 대장암 발생률이 높았으며, 운동량은 고활동 군에서 대장암 발생률이 높았다. 그 외에 크론병을 진단 받은 경우와 분변잠혈검사 음성인 경우 대장암 발생률이 높았다.

<표 4-12> 대장암 검진 대상자의 대장암 발생에 따른 기본적 특성

	전체	대장암				p-value
		미발생		발생		
	N	N	%	N	%	
나이						
50-59	824,783	812,392	98.5	12,391	1.5	
60-69	643,516	626,655	97.38	16,861	2.62	
70-79	246,306	237,989	96.62	8,317	3.38	
80 이상	23,028	22,239	96.57	789	3.43	
성별						
Male	767,661	745,519	97.12	22,142	2.88	<.0001
Female	969,972	953,756	98.33	16,216	1.67	
소득						
의료급여	4,546	4,422	97.27	124	2.73	<.0001
1분위	428,315	419,018	97.83	9,297	2.17	
2분위	436,740	427,428	97.87	9,312	2.13	
3분위	331,124	323,813	97.79	7,311	2.21	
4분위	536,908	524,594	97.71	12,314	2.29	
고혈압						
정상	408,840	401,184	98.13	7,656	1.87	<.0001
전기 고혈압	627,228	614,006	97.89	13,222	2.11	
고혈압	701,565	684,085	97.51	17,480	2.49	
당뇨병						
정상	1,336,838	1,309,532	97.96	27,306	2.04	<.0001
당뇨병	166,460	162,238	97.46	4,222	2.54	
당뇨병	234,335	227,505	97.09	6,830	2.91	
Total cholesterol						
평균 ± 표준편차	201.35 ± 34.79	201.4 ± 34.79		198.9 ± 34.89		
TG						
평균 ± 표준편차	133.17 ± 69.88	133.1 ± 69.85		137.1 ± 71.18		
HDL						
평균 ± 표준편차	54 ± 12.72	54.02 ± 12.72		53.22 ± 12.70		
LDL						
평균 ± 표준편차	120.42 ± 32.33	120.5 ± 32.32		118 ± 32.45		
허리둘레						
평균 ± 표준편차	61.62 ± 8.95	61.59 ± 8.95		62.87 ± 9.10		
고콜레스테롤						
정상	1,104,594	1,080,654	97.83	23,940	2.17	<.0001
고콜레스테롤	633,039	618,621	97.72	14,418	2.28	
BMI						
낮음	13,724	13,380	97.49	344	2.51	<.0001
정상	577,434	565,020	97.85	12,414	2.15	
비만 1단계	500,269	489,398	97.83	10,871	2.17	
비만 2단계	607,283	593,427	97.72	13,856	2.28	
비만 3단계	38,923	38,050	97.76	873	2.24	
흡연						
비흡연	1,209,650	1,186,457	98.08	23,193	1.92	<.0001
과거 흡연	280,436	272,233	97.07	8,203	2.93	
흡연	247,547	240,585	97.19	6,962	2.81	
음주						
비음주	1,167,770	1,144,447	98	23,323	2.00	<.0001
저위험 음주	492,172	479,546	97.43	12,626	2.57	
중위험 음주	39,713	38,547	97.06	1,166	2.94	
고위험 음주	37,978	36,735	96.73	1,243	3.27	
운동						
고활동군	191,277	186,662	97.59	4,615	2.41	<.0001
중활동군	585,837	572,737	97.76	13,100	2.24	
저활동군	960,519	939,876	97.85	20,643	2.15	
궤양성 대장염						
정상	1,735,225	1,696,919	97.79	38,306	2.21	0.9274
진단	2,408	2,356	97.84	52	2.16	
크론병						
정상	1,737,215	1,698,869	97.79	38,346	2.21	0.4492
진단	418	406	97.13	12	2.87	
분변잠혈검사 양성성						
양성성	1,635,178	1,601,292	97.93	33,886	2.07	<.0001
음성성	102,455	97,983	95.64	4,472	4.36	
총합	1,737,633	1,699,275	97.79	38,358	2.21	

(2) 대장암검진자의 cox-proportional hazards model 구축

예측모델에서는 빈도가 너무 작아서 불안정한 변수는 제외하였다. 예측모델에서 나이에 따라 가장 큰 격차를 보이는 것을 알 수 있으나 50세 이상의 대상에서는 일반 검진보다 그 격차가 줄어들었다. 그 외에 대부분의 변수들이 유의하였고 그 중 당뇨병, 음주, 분변잠혈검사에서 가장 큰 차이를 보였다.

<표 4-13> 대장암 검진 대상자의 대장암 발생 예측모델 (생존분석)

	HR	95% CI		p-value
		Low	High	
나이				
50-59	1.000	-	-	
60-69	1.700	1.666	1.734	<0.0001
70-79	2.365	2.323	2.407	<0.0001
80 이상	2.779	2.701	2.858	<0.0001
성별				
Male	1.000	-	-	
Female	0.700	0.665	0.735	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.005	0.966	1.044	0.7936
2분위	1.030	0.994	1.067	0.1239
1분위	1.015	0.975	1.054	0.4411
의료급여	1.174	0.954	1.393	0.1168
고혈압				
정상	1.000	-	-	
전기고혈압	0.998	0.963	1.032	0.9287
고혈압	1.041	1.008	1.074	0.0194
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.137	1.093	1.181	<.0001
당뇨병	1.257	1.230	1.284	<.0001
허리둘레				
1cm당	1.004	1.002	1.005	<.0001
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	1.018	0.996	1.040	0.055
흡연				
비흡연	1.000	-	-	
과거흡연	1.106	1.069	1.142	<.0001
흡연	1.166	1.128	1.204	<.0001
음주				
비음주군	1.000	-	-	
저위험 음주군	1.088	1.059	1.118	<.0001
중위험 음주군	1.242	1.171	1.314	<.0001
고위험 음주군	1.277	1.199	1.355	<.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	1.020	0.982	1.058	0.2575
저활동군	1.026	0.988	1.063	0.1575
분변잠혈검사				
양성	2.052	2.021	2.083	<.0001
음성	1.000	-	-	

(3) 대장암검진 대상자의 예측모델 (Logistic regression)

생존분석과 비슷한 결과가 나타났다.

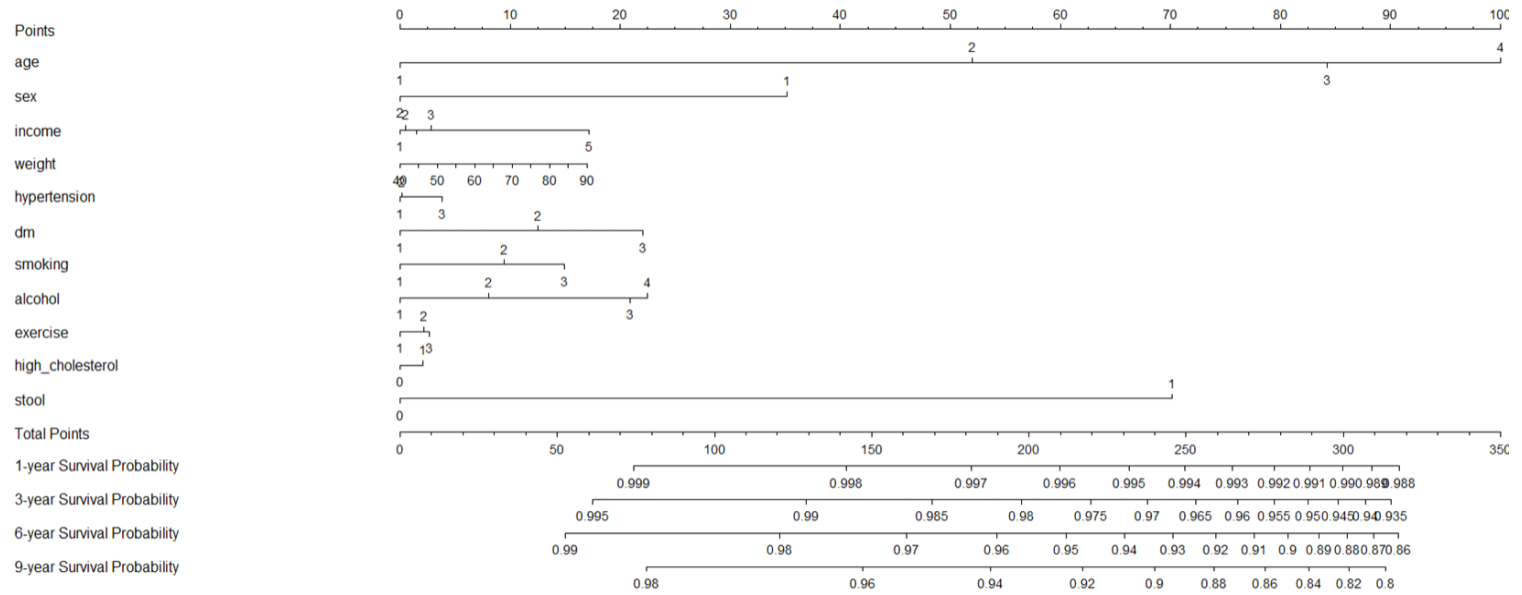
<표 4-14> 대장암 검진 대상자의 대장암 발생 예측모델 (로지스틱 회귀분석)

	OR	95% CI		p-value
		Low	High	
나이				
	50-59	1.000	-	-
	60-69	1.694	1.668	1.720
	70-79	2.267	2.229	2.305
	80 이상	2.247	2.156	2.338
성별				
	Male	1.000	-	-
	Female	0.723	0.683	0.763
소득				
	4분위	1.000	-	-
	3분위	1.008	0.965	1.050
	2분위	1.028	0.995	1.062
	1분위	1.016	0.982	1.050
	의료급여	1.121	0.902	1.340
고혈압				
	정상	1.000	-	-
	전기고혈압	0.998	0.960	1.037
	고혈압	1.035	1.001	1.068
당뇨병				
	정상	1.000	-	-
	당뇨병	1.137	1.090	1.184
	당뇨병	1.230	1.196	1.264
허리둘레				
	1cm당	1.005	1.003	1.006
고콜레스테롤				
	정상	1.000	-	-
	고콜레스테롤	1.020	0.992	1.048
흡연				
	비흡연	1.000	-	-
	과거흡연	1.091	1.046	1.135
	흡연	1.140	1.097	1.184
음주				
	비음주군	1.000	-	-
	저위험음주군	1.099	1.068	1.131
	중위험음주군	1.258	1.189	1.327
	고위험음주군	1.269	1.212	1.326
운동				
	고활동군	1.000	-	-
	중활동군	1.027	0.989	1.066
	저활동군	1.024	0.983	1.065
분변잠혈검사				
	양성	2.050	2.010	2.091
	음성	1.000	-	-

(4) 대장암검진 대상자의 예측모델 Nomogram

각각의 변수를 위로 선을 그으면 포인트를 알 수 있고 해당 포인트들을 전부 합친 것을 Total point에 대입하여 그 아래로 선을 그으면 1,3,6,9년 생존율을 알 수 있다.

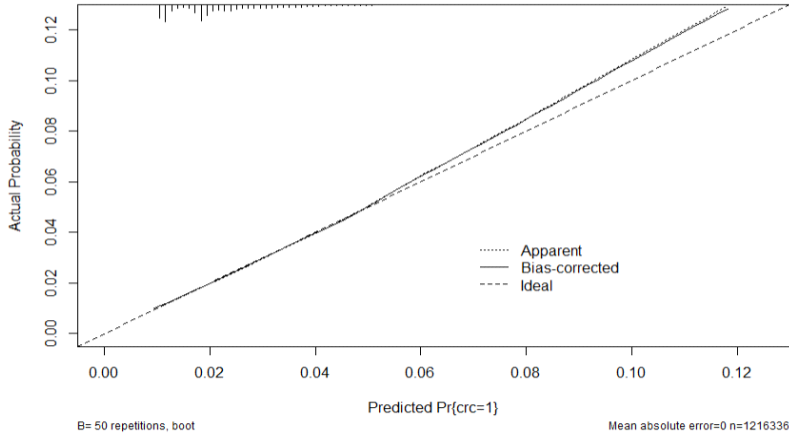
일반검진 대상자에 비해 다른 위험요소들의 점수가 전체적으로 상승하였으며, 분변잠혈검사의 점수가 높은 것으로 나타났다.



[그림 4-17] 대장암 검진 대상자의 대장암 발생 예측모델 Nomogram

(5) 대장암검진 대상자의 예측모델 Calibration plot

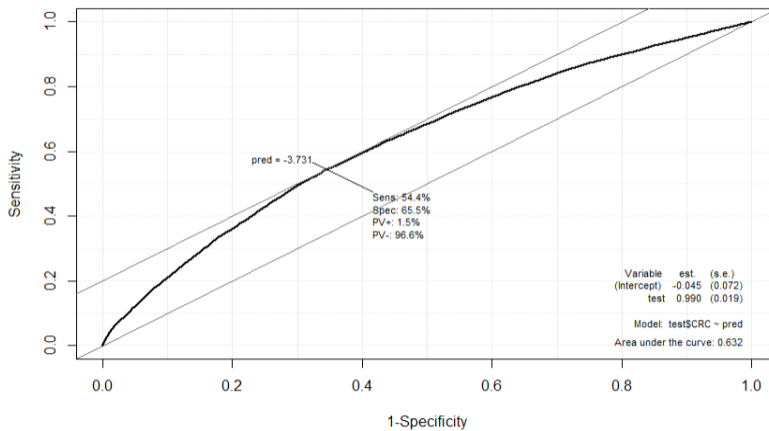
예측한 간암과 실제 관측된 간암이 얼마나 다른지 본 그래프이며 45도의 사선 점선에 가까울수록 예측과 실제 관측 값이 같은 이상적인 그래프이다. 이 그래프에서는 확률이 높은 부분에서 예측확률보다 실제 관측 값이 약간 높아지는 것을 알 수 있다.



[그림 4-18] 대장암 검진 대상자의 대장암 발생 예측모델 Calibration plot

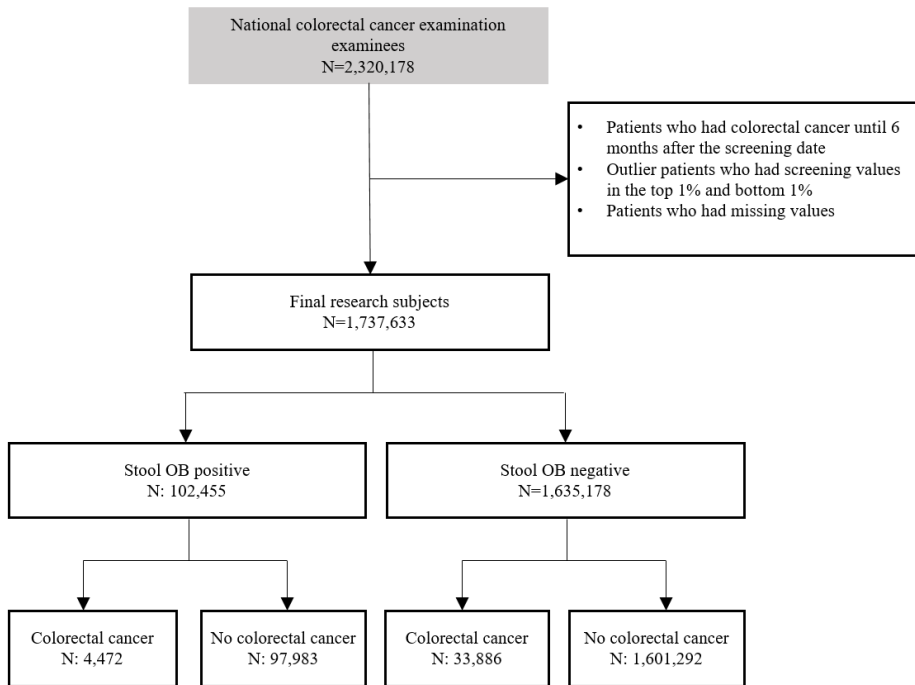
(6) 간암검진 대상자의 예측모델 ROC curve

본 예측모델의 AUC값은 0.632로 분석되었다.



[그림 4-19] 대장암 검진 대상자의 대장암 발생 예측모델 ROC curve





[그림 4-20] 대장암 검진 대상자의 대장암 발생 flow chart

### 3. 분변잠혈검사 양성 대상자 대장암 발생 예측

#### (1) 분변잠혈검사 양성 대상자의 일반적 특성

최대 10년의 추적관찰 기간 중에 전체 대상자 중 4.36%인 4,472명에서 대장암이 발생하였다. 80대에서 발생률이 가장 높았고 여성보다 남성에서 더 발생하였다. 경제 수준은 의료 급여에서 발생률이 가장 높게 나타났다. 고혈압, 당뇨병이 있는 경우 대장암 발생률이 높았다. 콜레스테롤 지표는 Total cholesterol, HDL, LDL 모두 대장암 발생 군에서 낮게 측정되었고, TG는 대장암 발생군에서 높게 측정되었다.

생활습관의 경우 과거 흡연 경험이 있는 군에서 대장암이 더 많이 발생하였다. 음주에서는 고위험 음주력이 있는 경우 대장암 발생률이 높았으며, 운동량은 고활동 군에서 대장암 발생률이 높았다. 그 외에 내시경 검사와 이중조영검사 미수검자인 경우 대장암 발생률이 높았다.

<표 4-15> 분변잠혈검사 양성 대상자의 대장암 발생에 따른 기본적 특성

	전체	대장암				p-value
		미발생		발생		
	N	N	%	N	%	
나이						
50-59	46,322	45,073	97.3	1,249	2.7	<.0001
60-69	37,846	35,956	95.01	1,890	4.99	
70-79	16,451	15,272	92.83	1,179	7.17	
80 이상	1,836	1,682	91.61	154	8.39	
성별						
Male	53,575	50,650	94.54	2,925	5.46	<.0001
Female	48,880	47,333	96.84	1,547	3.16	
소득						
의료급여	296	279	94.26	17	5.74	0.6803
1분위	25,565	24,429	95.56	1,136	4.44	
2분위	26,013	24,881	95.65	1,132	4.35	
3분위	19,441	18,587	95.61	854	4.39	
4분위	31,140	29,807	95.72	1,333	4.28	
고혈압	0					
정상	21,132	20,419	96.63	713	3.37	<.0001
전기고혈압	36,921	35,398	95.87	1,523	4.13	
고혈압	44,402	42,166	94.96	2,236	5.04	
당뇨병	0					
정상	76,741	73,639	95.96	3,102	4.04	<.0001
공복혈당장애	10,688	10,143	94.90	545	5.10	
당뇨병	15,026	14,201	94.51	825	5.49	
Total cholesterol						
평균 ± 표준편차	201.05 ± 35.26	201.2 ± 35.27		198.6 ± 34.97		<.0001
TG						
평균 ± 표준편차	135.72 ± 71.27	135.5 ± 71.14		141.1 ± 73.88		<.0001
HDL						
평균 ± 표준편차	54.04 ± 13.12	54.09 ± 13.13		52.97 ± 12.98		<.0001
LDL						
평균 ± 표준편차	119.56 ± 32.86	119.7 ± 32.87		117 ± 32.66		<.0001
허리둘레						
평균 ± 표준편차	62.45 ± 9.15	62.42 ± 9.15		63.28 ± 9.21		<.0001
고콜레스테롤						
정상	63,601	60,897	95.75	2,704	4.25	0.0241
고콜레스테롤	38,854	37,086	95.45	1,768	4.55	
BMI						
낮음	870	837	96.21	33	3.79	0.7155
정상	33,280	31,795	95.54	1,485	4.46	
비만1단계	29,118	27,873	95.72	1,245	4.28	
비만2단계	36,855	35,251	95.65	1,604	4.35	
비만3단계	2,332	2,227	95.50	105	4.50	
흡연						
비흡연	66,095	63,626	96.26	2,469	3.74	<.0001
과거흡연	18,015	16,976	94.23	1,039	5.77	
흡연	18,345	17,381	94.75	964	5.25	

	전체	대장암				p-value
		미발생		발생		
	N	N	%	N	%	
음주						
비음주군	64,251	61,793	96.17	2,458	3.83	<.0001
저위험 음주군	31,746	30,116	94.87	1,630	5.13	
중위험 음주군	3,094	2,923	94.47	171	5.53	
고위험 음주군	3,364	3,151	93.67	213	6.33	
운동						
고활동군	11,114	10,575	95.15	539	4.85	0.0166
중등도 활동군	34,201	32,696	95.6	1,505	4.40	
저활동군	57,140	54,712	95.75	2,428	4.25	
궤양성 대장염						
정상	102,099	97,640	95.63	4,459	4.37	0.5962
진단	356	343	96.35	13	3.65	
크론병						
정상	102,423	97,953	95.64	4,470	4.36	0.6498
진단	32	30	93.75	2	6.25	
내시경 여부						
미수검	75,626	71,564	94.62	4,065	5.38	<.0001
수검	26,829	26,419	98.48	407	1.52	
이중조영검사 여부						
미수검	97,030	92,764	95.6	4,270	4.40	<.0001
수검	5,425	5,219	96.28	202	3.72	
총합	102,455	97,983	95.64	4,472	4.36	

(2) 분변잠혈검사 양성 대상자의 cox-proportional hazards model 구축

예측모델에서는 빈도가 너무 작아서 불안정한 변수는 제외하였다. 예측모델에서 나이에 따라 가장 큰 격차를 보였으며, 그 외에 음주, 소득에서 가장 큰 차이를 보였다.

<표 4-16> 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 (생존분석)

	HR	95% CI		p-value
		Low	High	
나이				
50-59	1.000	-	-	
60-69	1.830	1.732	1.928	<0.0001
70-79	2.993	2.883	3.102	<0.0001
80 이상	3.730	3.502	3.957	<0.0001
성별				
Male	1.000	-	-	
Female	0.717	0.613	0.821	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.213	1.104	1.322	0.0011
2분위	1.239	1.159	1.318	<0.0001
1분위	1.182	1.093	1.270	0.0005
의료급여	1.479	0.830	2.128	0.1912

	HR	95% CI		p-value
		Low	High	
고혈압				
정상	1.000	-	-	
전기고혈압	1.066	0.970	1.162	0.1575
고혈압	1.127	1.042	1.212	0.0049
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.100	1.003	1.197	0.0529
당뇨병	1.119	1.024	1.215	0.0192
허리둘레				
1cm당	1.001	0.996	1.006	0.7156
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	1.078	1.012	1.145	0.023
흡연				
비흡연	1.000	-	-	
과거흡연	1.240	1.154	1.327	<0.0001
흡연	1.159	1.041	1.276	0.0112
음주				
비음주군	1.000	-	-	
저위험 음주군	1.198	1.117	1.279	<0.0001
중위험 음주군	1.252	1.049	1.455	0.0346
고위험 음주군	1.463	1.302	1.623	<0.0001
운동				
고활동군	1.000	-	-	
중등도 활동군	0.970	0.830	1.111	0.7014
저활동군	0.943	0.811	1.075	0.3977

(3) 분변잠혈검사 양성 대상자의 예측모델 (Logistic regression)

생존분석과 비슷한 결과가 나타났다.

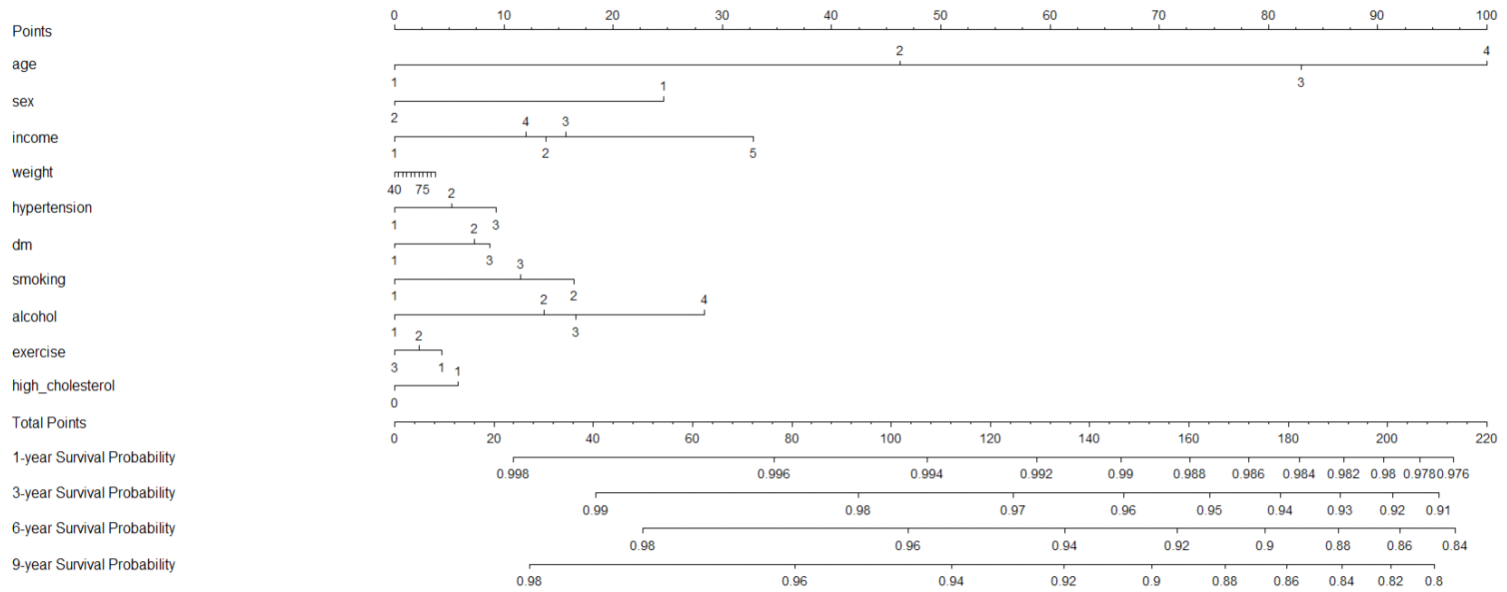
<표 4-17> 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 (로지스틱 회귀분석)

	OR	95% CI		p-value
		Low	High	
나이				
50-59	1.000	-	-	
60-69	1.843	1.747	1.940	<0.0001
70-79	2.875	2.773	2.978	<0.0001
80 이상	3.094	2.841	3.346	<0.0001
성별				
Male	1.000	-	-	
Female	0.746	0.629	0.863	<0.0001
소득				
4분위	1.000	-	-	
3분위	1.202	1.101	1.304	0.0004
2분위	1.235	1.132	1.339	<0.0001
1분위	1.178	1.090	1.266	0.0005
의료급여	1.390	0.870	1.910	0.1762
고혈압				
정상	1.000	-	-	
전기고혈압	1.074	0.947	1.202	0.3035
고혈압	1.139	1.003	1.275	0.0828

	OR	95% CI		p-value
		Low	High	
당뇨병				
정상	1.000	-	-	
공복혈당장애	1.091	0.993	1.189	0.0715
당뇨병	1.105	1.003	1.207	0.0615
허리둘레				
1cm당	1.002	0.997	1.007	0.3347
고콜레스테롤				
정상	1.000	-	-	
고콜레스테롤	1.090	1.007	1.174	0.066
흡연				
비흡연	1.000	-	-	
과거흡연	1.251	1.136	1.365	0.0002
흡연	1.149	1.040	1.258	0.0146
음주				
비음주군	1.000	-	-	
저위험 음주군	1.205	1.110	1.299	<0.0001
중위험 음주군	1.255	1.045	1.465	0.0317
고위험 음주군	1.426	1.229	1.623	0.0002
활동				
고활동군	1.000	-	-	
중등도 활동군	0.994	0.877	1.110	0.6855
저활동군	0.957	0.837	1.076	0.3274

(4) 분변잠혈검사 양성 대상자의 예측모델 Nomogram

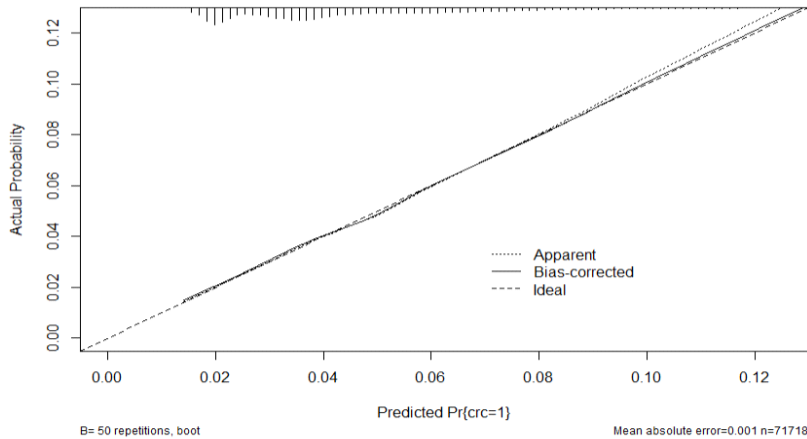
각각의 변수를 위로 선을 그으면 포인트를 알 수 있고 해당 포인트들을 전부 합친 것을 Total point에 대입하여 그 아래로 선을 그으면 1,3,6,9년 생존율을 알 수 있다. 음주, 소득, 나이에서 큰 점수를 보였다



[그림 4-21] 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 Nomogram

(5) 분변잠혈검사 양성 대상자의 예측모델 Calibration plot

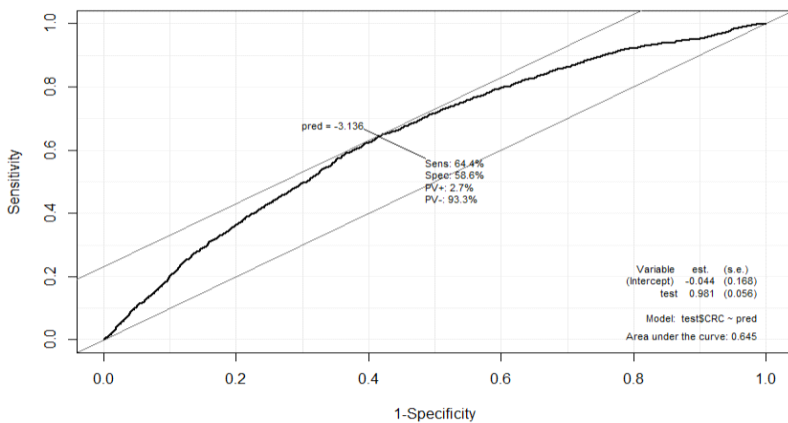
예측한 대장암과 실제 관측된 대장암이 얼마나 다른지 본 그래프이며 45도의 사선 점선에 가까울수록 예측과 실제 관측 값이 같은 이상적인 그래프이다. 이 그래프에서는 확률이 높은 부분에서 예측확률보다 실제 관측 값이 약간 높아지는 것을 알 수 있다.



[그림 4-22] 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 Calibration plot

(6) 분변잠혈검사 양성 대상자의 예측모델 ROC curve

본 예측모델의 AUC값은 0.645로 분석되었다.



[그림 4-23] 분변잠혈검사 양성 대상자의 대장암 발생 예측모델 ROC curve

(7) 분변잠혈검사 양성 대상자 중 내시경/대장이중조영검사자 검사결과

분변잠혈검사 양성 대상자의 내시경 검사 결과 대장용종이 45.65%로 가장 많았으며 이상 병변의 위치는 표와 같다.

<표 4-18> 내시경 검사 결과

	N	%
내시경 검사 결과		
정상	9,092	33.89
대장용종	12,248	45.65
기타	5,489	20.46
내시경 이상병변 위치		
회장 말단부	356	1.3
맹장	1,091	4.1
상행 결장	3,537	13.2
간 만곡	995	3.7
횡행 결장	2,541	9.5
비 만곡	265	1.0
하행 결장	1,937	7.2
에스 결장	4,267	15.9
직장	3,512	13.1
항문	2,739	10.2
기타	5,589	20.8
총합	26,829	100

분변잠혈검사 양성 대상자의 대장이중조영 검사 결과 정상이 74.29%로 가장 많았으며 이상 병변의 위치는 표와 같다.

<표 4-19> 대장이중조영검사 결과

	N	%
대장이중조영검사 결과		
정상	4,030	74.29
대장용종	319	5.88
기타	1,076	19.83
대장이중조영검사 이상병변 위치		
회장 말단부	31	0.6
맹장	139	2.6
상행 결장	436	8.0
간 만곡	35	0.6
횡행 결장	64	1.2
비 만곡	20	0.4
하행 결장	160	2.9
에스 결장	243	4.5
직장	127	2.3
항문	86	1.6
기타	4,084	75.3
총합	5,425	100



### 제3절 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과 연구

#### 1. 연구대상자 특성 파악

사후관리 프로그램 사용 전과 사용 이후의 연구대상자 특성을 살펴보았다. <표 4-20> 키, 몸무게, 허리둘레, 공복혈당, 중성지방, 고밀도콜레스테롤의 경우, 사후관리 프로그램 사용 전보다 후에 평균이 감소하였고 수축기혈압, 이완기혈압의 경우, 사후관리 프로그램 사용 전보다 후에 평균이 증가하였다.

<표 4-20> 연구대상자 특성

※변수 \ 사용 여부	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
키	164.80 ± 9.07	164.76 ± 9.03
몸무게	66.86 ± 13.36	66.72 ± 13.08
허리둘레	79.65 ± 10.08	79.34 ± 9.03
수축기혈압	123.74 ± 11.53	124.79 ± 10.22
이완기혈압	75.39 ± 9.18	75.49 ± 7.80
공복혈당	99.22 ± 13.20	97.52 ± 11.20
중성지방	129.68 ± 101.69	109.19 ± 55.00
고밀도콜레스테롤	46.57 ± 19.02	45.19 ± 18.09

※p-value ≥ 0.05 (NS), Paired T-Test

#### 2. 대사증후군 개선효과

##### (1) 대사증후군 타입 분포 비교

사후관리 프로그램 사용 전과 후의 대사증후군 위험요인 수를 분석하여 대사증후군 타입 분포를 살펴보았다. 대사증후군 타입 ‘위험군’은 32명(50.8%)에서 34명(54.0%)으로 늘었고 ‘대사증후군’은 18명(28.6%)에서 16명(25.4%)로 줄었다. <표 4-21>

<표 4-21> 대사증후군 타입 분포 비교

※대사증후군 타입	사후관리 프로그램 사용 전		사후관리 프로그램 사용 후	
	수	비율	수	비율
정상군	13	20.6%	13	20.6%
위험군	32	50.8%	34	54.0%
대사증후군	18	28.6%	16	25.4%

※p-value ≥ 0.05 (NS), Chi-square Test

(2) 대사나이 변화 분석

대사나이의 평균의 분석에서는 사후관리 프로그램 사용 전 평균은 2.20세, 사후관리 프로그램 사용 후 평균은 1.72세를 보였다. 사후관리 프로그램 사용 전 보다 사용 후에 대사나이 평균이 0.48세 감소하였다. <표 4-22>

사후관리 프로그램 사용 전과 후의 대사나이를 산출하여 대사나이 타입 분포를 살펴보았다. 대사나이 타입 ‘ 좋음 ’은 6명(9.5%)에서 8명(12.7%)으로 늘었고, 대사나이 타입 ‘ 평균 ’은 30명(47.6%)에서 39명(61.9%)으로 늘었으며 대사나이 타입 ‘ 나쁨 ’은 27명(42.9%)에서 16명(25.4%)으로 줄었다. <표 4-23>

<표 4-22> 대사나이 평균 비교

	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
※대사나이 평균	2.20	1.72

※p-value ≥ 0.05 (NS), Paired T-Test

<표 4-23> 대사나이 타입 분포 비교

※대사나이 타입	사후관리 프로그램 사용 전		사후관리 프로그램 사용 후	
	수	비율	수	비율
좋음 (대사나이 ≤ -1SD)	6	9.5%	8	12.7%
평균 (-1SD < 대사나이 < +1SD)	30	47.6%	39	61.9%
나쁨 (+1SD ≤ 대사나이)	27	42.9%	16	25.4%

※p-value ≥ 0.05 (NS), Chi-square Test

### 3. 생활습관 개선효과

#### (1) 설문지 분석

운동 횟수 평균을 비교해 보았을 때, 사후관리 프로그램 사용 전 평균은 4.11회, 사후관리 프로그램 사용 후 평균은 4.38회를 보였다. 사후관리 프로그램 사용 전 보다 사용 후에 운동 횟수 평균이 0.27회 증가하였다. <표 4-24>

사후관리 프로그램 사용 전과 후의 흡연자 분포를 살펴보았다. 비흡연자는 58명(92.1%)에서 59명(93.7%)로 늘었고, 흡연자는 5명(7.9%)에서 4명(6.3%)으로 줄었다. <표 4-25>

사후관리 프로그램 사용 전과 후의 음주자 분포를 살펴보았다. 음주를 하지 않는 사람은 38명(60.3%)에서 32명(50.8%)로 줄었고, 음주자는 25명(39.7%)에서 31명(49.2%)으로 늘었다. <표 4-26> 음주자의 음주 횟수를 비교해 보았을 때, 사후관리 프로그램 사용 전 평균은 1.8회, 사후관리 프로그램 사용 후 평균은 1.5회를 보였다. 사후관리 프로그램 사용 전 보다 사용 후에 음주 횟수 평균이 0.3회 감소하였다. <표 4-27> 음주자의 음주량(잔)을 비교해 보았을 때, 사후관리 프로그램 사용 전 평균은 5.52잔, 사후관리 프로그램 사용 후 평균은 4.35잔을 보였다. 사후관리 프로그램 사용 전 보다 사용 후에 음주량(잔)의 평균이 1.17잔 감소하였다. <표 4-28>

사후관리 프로그램 사용 전과 후의 수면시간 분포를 살펴보았다. 수면시간 7시간 이하 혹은 8시간 이상 해당자의 경우, 47명(74.6%)에서 37명(58.7%)로 줄었고, 수면시간 7-8시간 해당자는 16명(25.4%)에서 26명(41.3%)으로 늘었다. <표 4-29>

스트레스 점수 평균을 비교해 보았을 때, 사후관리 프로그램 사용 전 평균은 3.04점, 사후관리 프로그램 사용 후 평균은 3.03점을 보였다. 사후관리 프로그램 사용 전 보다 사용 후에 스트레스 점수 평균이 0.01점 감소하였다. <표 4-30>

사후관리 프로그램 사용 전과 후의 영양 습관 분포를 살펴보았다. 비만, 혈압, 고지혈증, 당뇨 등 4가지 항목의 영양 습관이 좋지 않은 사람 평균이 28명(44.44%)에서 26.75명(42.46%)으로 줄었다. <표 4-31>

<표 4-24> 운동 횟수 평균 비교

	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
※운동 횟수 평균	4.11	4.38

※p-value ≥ 0.05 (NS), Paired T-Test

<표 4-25> 흡연자 분포 비교

※흡연 여부	사후관리 프로그램 사용 전		사후관리 프로그램 사용 후	
	수	비율	수	비율
비흡연	58	92.1%	59	93.7%
흡연	5	7.9%	4	6.3%

※p-value ≥ 0.05 (NS), Chi-square Test

<표 4-26> 음주자 분포 비교

※음주 여부	사후관리 프로그램 사용 전		사후관리 프로그램 사용 후	
	수	비율	수	비율
음주 X	38	60.3%	32	50.8%
음주 O	25	39.7%	31	49.2%

※p-value ≥ 0.05 (NS), Chi-square Test

<표 4-27> 음주 횟수 평균 비교

	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
※음주 횟수 평균	1.8	1.5

※p-value ≥ 0.05 (NS), Paired T-Test

<표 4-28> 음주량(잔) 평균 비교

	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
※음주량(잔) 평균	5.52	4.35

※p-value ≥ 0.05 (NS), Paired T-Test

<표 4-29> 수면시간 분포 비교

※수면시간	사후관리 프로그램 사용 전		사후관리 프로그램 사용 후	
	수	비율	수	비율
7시간 이하 혹은 8시간 이상	47	74.6%	37	58.7%
7-8시간	16	25.4%	26	41.3%

※p-value ≥ 0.05 (NS), Chi-square Test

<표 4-30> 스트레스 점수 평균 비교

	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
※스트레스 점수 평균	3.04	3.03

※p-value ≥ 0.05 (NS), Paired T-Test

<표 4-31> 영양 습관 분포 비교

영양 관련 습관		사후관리 프로그램 사용 전		사후관리 프로그램 사용 후	
		수	비율	수	비율
영양1 (비만)	1(예)	38	60.32%	37	58.73%
	2(아니오)	25	39.68%	26	41.27%
영양2 (혈압)	1(예)	35	55.56%	36	57.14%
	2(아니오)	28	44.44%	27	42.86%
영양3 (고지혈증)	1(예)	19	30.16%	21	33.33%
	2(아니오)	44	69.84%	42	66.67%
영양4 (당뇨)	1(예)	20	31.75%	13	20.63%
	2(아니오)	43	68.25%	50	79.37%
평균	1(예)	28	44.44%	26.75	42.46%
	2(아니오)	35	55.56%	36.25	57.54%

(2) 걸음 수 변화량 분석

걸음 수 평균을 비교해 보았을 때, 사후관리 프로그램 사용 전 평균은 6,305보, 사후관리 프로그램 사용 후 평균은 6,628보를 보였다. 사후관리 프로그램 사용 전 보다 사용 후에 걸음 수 평균이 323보 증가하였다. <표 4-32>

<표 4-32> 걸음 수 평균 비교

	사후관리 프로그램 사용 전	사후관리 프로그램 사용 후
※걸음 수 평균	6,305	6,628

※p-value ≥ 0.05 (NS), Paired T-Test

## 제4절 건강in 에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고

건강in은 인터넷상 검증되지 않고 무분별하게 서비스 되고 있는 건강정보로부터 국민 건강을 보호하기 위하여 전문가가 검증한 신뢰성 있는 건강정보를 제공하는 건강정보 전문사이트이다. 다양한 분야의 다양한 건강정보 제공과 함께 국가건강검진자료를 이용하여 개인의 위험도 평가 후 건강개선자료를 제공, 건강나이 알아보기, 비만·흡연자 및 대사성질환자를 위한 맞춤형 생활습관 개선 프로그램 정보 등 개인별 맞춤형 건강정보를 제공하고 있다.

대국민 교육 자료로는 질병관리본부에서 관리하고 있는 국가건강정보포털이 있다. 국민의 수요에 기반하여 검증된 양질의 건강, 질병 정보를 통합적·체계적으로 제공하는 공공포털이다. 검증된 정보(의학전문가의 광범위한 검증을 통해 신뢰할 수 있는 양질의 건강정보를 제공), 알기 쉬운 정보(쉬운 용어 사용, 시각자료 활용 등을 통해 일반 국민이 이해하기 쉬운 건강정보 제공), 통합적 정보(질병정보, 의료기관 정보, 의약품 정보 등 건강관련 정보의 통합·연계)로 가고 있다. 국가건강정보 포털의 서비스 콘텐츠 현황은 현재 총 618건으로 질병 관련된 콘텐츠가 358개로 가장 많다.

환자에 대한 기록으로는 각 의료기관의 정보와 함께 진료정보교류사업을 통한 EHR Data 가 수집되고 있다. 진료정보교류사업은 개인 정보 제공에 동의한 진료기록을 국민이 원하는 곳(의료기관)에 전자적인 방식(안전성·정확성)으로 전송하여 의사가 진료에 참고(신속성·연속성) 할 수 있도록 제공하는 서비스로 추진되고 있다. 2016년에 의료법이 개정되어 법적 근거가 마련되었고 2017년에 제도가 갖춰지고 본 사업을 할 수 있는 기반이 마련되었다. 2018, 2019년 진료정보교류사업 추진을 위한 거점의료기관과 문서 저장소가 전국적으로 구축이 되어 네트워크가 완성되었다. 14개의 거점 병원의 문서저장소와 1개의 공공 저장소를 구축하였다. 52개의 거점 의료기관이 있으며 거점 41개, 단위 기관 11개의 의료기관이 있으며 6,439개 병의원이 참여하고 있다. 4년 동안 11.2만 명의 환자가 21.4만 건의 진료정보교류서비스를 이용하였고 매년 200% 이상 진료정보교류 실적 증가가 있어 왔다. 교류되는 Data로는 진료 의뢰시에는 진료의뢰서, 진료기록(진단, 검사, 약물, 수술, 알리지 및 영상 등)이 제공되고, 의료기관에 내원한 환자의 진료기록은 요약지 형태로 작성하여 문서 저장소에 보관하여 제공한다. 영상 정보 교류는 CT,

MRI 등 영상정보파일과 판독 소견서를 제공한다. 진료 회송 시에는 진료회송서, 진료기록(진단, 검사, 약물, 수술, 알리지 및 영상 등) 등이 제공된다. 교류 data의 문제점은 먼저 정형 data도 품질이 좋지 않다. 병원의 20%의 정형 data 중에서도 청구 목적을 위한 10 %정도의 정보만 이용 가능하다. 또한 80%가 코드화 되어 있지 않은 경과일지, 영상 기록지 등과 같은 비정형 data라는 것이다. 비정형 data를 어떻게 coding 할 것인가가 중요한 issue 가 되고 있다.

국가에서 주도하고 있는 4차산업혁명위원회 디지털헬스특위에서는 예방·수요자 중심의 개인 의료데이터 연계·활용을 통한 국민건강증진 및 의료서비스 혁신, 그리고 이를 통한 의료체계 목표(환자중심 의료, 의료의 질 향상, 의료비 증가 억제) 달성을 목표로 의료데이터 이용 활성화를 방향으로 설정하고 있다. 개인 주도형으로 의료데이터 활용을 위하여 어디 한군데에 정보를 모으지 않고 개인 의료데이터 보호·보안 및 개인 주도적 접근성·활용이 보장된 My Healthway라는 gate way를 구축하여 의료기관, 보험, 고용정보원, 산재 등 여러 가지 종류의 정보가 개인사용 동의를 하면 활용이 될 수 있도록 중심적인 역할을 할 수 있고, 정보가 필요할 때 언제든지 활용할 수 있도록 구축을 준비해 가고 있다. 개인 주도형 의료데이터 중점 과제로는 개인 중심 의료데이터 통합·연계 체계 구축, 개인 정보 동의 및 보호 체계 구축, 개인중심 의료데이터 기반 서비스 발굴, 의료기관 및 국민 참여 활성화, 지속 가능한 혁신을 위한 거버넌스 등이 제시되고 있다.

전 세계적으로 PHR은 정책적으로 중요하며, real time으로 system에 접속해 나아가는 방향이다. 공단 data가 국민에게 건강증진, 건강개선 서비스를 고도화하여 제공하기 위해서는 목표를 세분화해서 연령, 지역, 질병, risk factor 별로 개발이 필요할 것이며, 현재 시점에서 할 수 있는 서비스가 무엇인지 define 하는 것이 필요하다. 국민건강보험 공단의 건강in은 국가적 인프라로 평균적·보편적·대표적 PHR로써 기능하고, 기업은 창의적인 접근을 통하여 국민건강증진에 기여하도록 역할의 세분화도 필요하다. 공단에서는 국가 검진을 통한 한국인 평균 data와 개인의 data를 비교하여 보여줌으로써 어떤 건강행동을 해야 다른 disease를 줄일 수 있을 지에 대한 정보를 제공하여 개인에게 특화된 건강정보 서비스를 제공할 수 있을 것이다. 사망과 주요 질환 (당뇨, 비만, 심질환, 뇌졸중, 암 등) 이외에도 노화 관련 질병 의 질병 예측서비스도 국민 개개인에게 매우 중요하다. 기타 공공 기관의 data도 건강인의 목표에 맞추어 결합한 서비스의 제공이 필요하다.

# 제5장

## 결론 및 제언

제1절 결 론	97
제2절 제 언	99

---





# 제5장 결론 및 제언

## 제1절 결론

### 1. 간암 질환 예측 연구

일반검진자 대상으로 전체 대상자의 1.7%인 150,922명에서 간암 발생하였다. HBV 감염자에서 간암 발생률이 가장 높았다. 그 외에 C형간염, 기타 간염, 간경변, 나이 등에서 큰 빈도 차이를 보였다. 일반검진 대상자로 간암 발생 예측 모델을 구축한 결과 B형간염 진단 여부에서 HR이 7.895 가장 높았다. 그 외에 고혈압과 운동을 제외한 대부분의 변수들이 유의하게 HR를 높였다. 모델의 검증력은 AUC curve 0.715로 확인되었다.

일반 검진자 중 간암검진자의 위험군으로 선정한 HBV, HCV, 간경화를 제거한 군을 기준으로 예측모델을 개발하였다. 나이가 가장 강력한 Hazard ratio를 나타냈다. 80세 이상은 HR이 5.361이었다. 흡연, 음주, 운동 등 생활습관이 악화될수록 HR를 증가하는 결과가 나타났다. 흡연하는 경우 HR는 1.157, 고위험 음주하는 경우 HR는 1.527으로 간암의 위험도가 올라갔다.

간암검진자의 일반적 특성을 분석하였다. 전체 수검자를 10년간 관찰한 결과 10.14%인 17,390명에서 간암 발생하였다. 간초음파 검사에서 이형결절이 진단된 경우 가장 간암발생률이 높았다. 그 외 거친에코상, C형간염, B형간염, 간경변, 음주, 흡연 등에서 큰 발생률 차이를 보였다. 간암검진 수검자를 대상으로 예측모형을 개발하였다. HR는 이형결절이 HR 3.072, B형간염이 HR 2.648, C형간염이 HR 2.116, AFP이 HR 1.568, 간경변이 HR 1.400으로 나타났다. 본 연구에서 대입한 Fatty Liver Index의 경우 HR가 1.009로 유의하게 나타났다.

### 2. 대장암 질환 예측 연구

일반수검자를 대상으로 위험인자 및 생활습관과 대장암과의 관련성을 분석한 결과

나이에 따라 가장 큰 격차를 보였고, 남성에서 위험도가 높게 나타났으며, 의료급여 대상자인 경우(HR 1.246), 고혈압(HR 1.068), 당뇨병(HR 1.233)이 있는 경우, 콜레스테롤 수치가 높은 경우(HR 1.026), 흡연자(HR 1.111), 고위험 음주군인 경우(HR 1.263), 저활동군인 경우(HR 1.024) 위험도가 높게 나타났다.

대장암검진 수검자를 대상으로 위험인자 및 생활습관과 대장암과의 관련성을 분석한 결과 나이에 따라 가장 큰 격차를 보였고, 남성에서 위험도가 높게 나타났으며, 의료급여 대상자인 경우(HR 1.174), 고혈압(HR 1.041), 당뇨병(HR 1.257)이 있는 경우, 콜레스테롤 수치가 높은 경우(HR 1.018), 흡연자(HR 1.166), 고위험 음주군인 경우(HR 1.277), 저활동군인 경우(HR 1.026), 분변잠혈검사 양성인 경우(HR 2.052) 위험도가 높게 나타났다.

분변잠혈검사 양성 수검자를 대상으로 위험인자 및 생활습관과 대장암과의 관련성을 분석한 결과 나이에 따라 가장 큰 격차를 보였고, 남성에서 위험도가 높게 나타났으며, 의료급여 대상자인 경우(HR 1.479), 고혈압(HR 1.127), 당뇨병(HR 1.119)이 있는 경우, 콜레스테롤 수치가 높은 경우(HR 1.078), 흡연자(HR 1.159), 고위험 음주군인 경우(HR 1.463) 위험도가 높게 나타났다.

### 3. 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과

총 63명의 연구대상자 연구에서 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선효과를 확인할 수 있었다.

대사증후군 개선효과의 측면으로는 사후관리 프로그램을 사용하기 전보다 후에 대사증후군 타입 비율이 3.2% 감소하였고, 대사나이 평균이 0.48세 감소하였으며 대사나이 타입 '나쁨' 비율이 17.5% 감소하였다.

생활습관 개선효과의 측면으로는 운동 횟수 증가, 흡연자 감소, 음주횟수의 감소와 음주량(잔) 감소가 나타났다. 수면시간의 경우, '7-8시간' 수면 해당자가 10명 증가하였고 스트레스 점수의 평균이 0.01점 감소하였다. 비만, 혈압, 고지혈증, 당뇨 등 4가지 항목의 영양 습관이 좋지 않은 사람이 평균치로 약 2명 감소하였다. 또한, 사후관리 프로그램을 사용하기 전보다 후에 걸음 수 평균도 증가하였다.

### 4. 건강in 에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고

PHR의 최종 정착지는 한 개인의 평생 건강기록 서비스를 제공하고, 태아~사망, 일상생활~병원을 연결하며 소아/청소년, 젊은 여성에 대한 접근을 용이하게 하는데 있다. 대국민 중심의 건강관리 도구로 활용될 수 있도록 국민 참여와 의사소통의 도구가 되어

야 할 것이며, 암, 당뇨, 만성 신질환과 같은 특정 질병 PHR로 활용하고 국민들의 자가 보고 결과와 개인의 진료정보교류를 위해 활용되어야 한다.

PHR 인터페이스를 위해 건강검진자료, 전자의무기록, 공공보건 서비스, 대국민 교육 자료, wearables, 디지털 치료제 등을 활용할 수 있을 것이다. 건강증진 서비스와 임상진료와의 통합을 통해 지불/비즈니스 모델을 구축하고, 필요시 119 정보를 연계하여 응급 상황에서의 PHR 서비스에 활용하고 공중보건 감시를 함으로써 PHR을 공중보건에도 활용할 수 있을 것이다. 정확성, 저장, 사용, 통합 및 교류 등을 통해 안전하고 신뢰 있는 PHR platform으로 나아가야 한다.

## 제2절 제언

### 1. 간암 질환 예측 연구

이 연구는 우리나라 국민 전체를 반영하는 최신 국민건강보험공단자료를 활용하여 일반검진대상자, 일반검진 대상자 중 간암위험도 높은 군 제외한 대상자, 간암검진 대상자 세 가지 군에 대해 분석을 하였다. 결론적으로 이는 기존 암예측 연구와 비교한 결과 HR가 비슷하여 상대적으로 의미 있는 모형이었다고 할 수 있겠다. 다만 B형간염의 HR가 7.895로 이전 연구에 비해 낮은 것은 2009년 검진 전과 2009년 검진 이후 6개월 이내에 간암코드가 있는 경우를 wash out 하면서 일부 대상자가 제외되었기 때문일 것으로 생각된다. AUC curve로 검증력을 판단한 결과 일반검진자 대상 0.715, 간암검진자 대상 0.708으로 나타났다.

특히 본 연구에서는 일반검진 대상자(약 878만 명) 중에 B형간염, C형간염, 간경변 등 사실상 '간암검진대상자'에 포함될 수 있었던 사람은 제외하고 예측 모델 구축하였다. 그 이유로 이미 high risk group인 사람을 제외하고 risk가 없다고 생각했는데 간암이 발생하는 경우를 targeting해서 예측모형을 만드는 것이 general population(간암에 대한 traditional risk factor가 없는 대다수의 사람들)에 대해 조언을 줄 수 있기 때문이다.

또한 본 연구에서는 혈청검사 결과 등 기본적인 수치로 추가적인 영상학적 검사 없이 사용할 수 있는 fatty liver index와 암의 상관성을 보았다. 본 연구 결과 AFI는 유의미하였으나 HR가 1.01에 그쳤다. AFI만으로 간암을 예측하는 데는 무리가 있었다.

추후 간암예측모형의 검증력을 높이기 위해 간암 발생과 관련될 수 있는 가족력 및 그 외 생략된 변수들을 도입하여 고도화된 모형을 구축할 필요가 있겠다. 본 연구의

한계점으로는 고위험군(간암검진 대상자)을 대상으로 예측모델을 만드는 것은 이미 HBV 항바이러스제 복용, virus titer, HBeAg status 등 공단 자료로 보충하기는 어려운 자료의 획득의 문제점이 있다.

## 2. 대장암 질환 예측 연구

일반검진, 대장암검진, 분변잠혈검사 양성 수검자에서 공통적으로 연령 증가에 따라, 남성인 경우, 고혈압, 당뇨병, 고콜레스테롤증에서 대장암 발생률과 위험도가 높은 결과를 보였다. 생활습관에서는 흡연자, 고위험 음주군에서 대장암 발생률과 위험도가 높은 것으로 나타났다. 이러한 위험군에 대하여 대장암 발생 예방을 위해 생활습관 관리가 필요할 것으로 생각된다.

대장암검진 수검자를 대상으로 한 예측모델에서 나이에 따라 격차를 보였으나 50세 이상의 대상에서는 일반검진보다 그 격차가 줄어들었다. 이는 분석 대상자를 일반검진 대상자는 40세 이상, 대장암 검진 대상자는 50세 이상으로 분류했기 때문인 것으로 보인다.

대장암검진 대상자의 대장암 발생 예측 모델의 AUC 값은 0.632로 분석되었다. 이는 이 연구의 한계점으로 대장암 가족력이나 식이와 같은 요인은 대장암 발생 예측에 중요하나 2009년 당시 검진 자료에는 이에 대한 data가 없어 포함시키지 못 하였으며, 추후 국가검진에 사용하는 문진표에 보다 자세한 데이터 수집이 필요할 것으로 사료된다.

분변잠혈검사 양성 대상자에서 대장내시경이나 이중조영검사를 받지 않은 대상자에서 대장암 발생이 더 높은 빈도로 나타나, 분변잠혈검사 양성인 수검자가 내시경 검사를 받는 비율을 높일 수 있는 방안이 필요한 것으로 보인다.

## 3. 국가검진대상자들의 사후관리 프로그램을 통한 대사증후군 개선 효과

약 3개월간 모집된 대상자를 분석하였기에 짧은 기간에 모집할 수 있는 인원수의 한계로 인하여 분석 대상자의 50.8%인 63명의 적은 인원으로 분석이 진행되었으며 통계적 유의성을 입증하는데 한계점이 있었다. 또한 COVID-19 상황으로 인하여 실외 활동이나, 병원 방문 기피 현상 등으로 인해 모든 참여 대상자들의 개선 효과를 비교해 보는 데는 어려움이 있었다. 이에 대해 연구 대상자 모집 기간 및 모집단 수를 늘려 통계적 유의성을 보완하는 연구 모형으로 추후 연구 진행이 필요하다. 또한 공단검진 시행 후, 검진 결과 제공과 더불어 대사증후군 위험요인이 있는 대상자들에게 고도화된 맞춤형 건강관리 프로그램의 꾸준한 제공과 함께, 대사증후군의 위험요인을 감소시킬 수 있는 더욱 고도화 연구가 필요할 것으로 판단된다.

#### 4. 건강in 에서의 개인별 맞춤형 통합 건강관리의 고도화 및 방향성 제고

국민건강보험공단의 건강in은 평균적·보편적·대표적 PHR로써 기능하도록 공공의 자료를 통합하여 나아가야 하며, 기업은 지속적·연계적인 PHR service를 제공하는 등 역할에 있어서 차이가 있다.

공단 data가 국민에게 제공되는 건강증진, 건강개선 서비스는 목표를 세분화해서 연령, 지역, 질병, risk factor 별로 개발이 필요할 것이며, 현재 시점에서 할 수 있는 서비스가 무엇인지 define 하는 것이 필요하다. 예를 들어 국가 검진을 통해 한국인 평균 data와 개인의 data를 비교하여 보여줌으로써 어떤 건강행동을 해야 다른 disease를 줄일 수 있을 지에 대한 정보를 제공하여 개인에게 특화된 건강정보 서비스를 제공할 수 있을 것이다.

사용할 data에 대해서는 개인별 맞춤형 자료를 계속 고도화해야 한다. 지금까지의 연구에서는 사망과 주요 질환 (당뇨, 비만, 심질환, 뇌졸중, 암 등)의 발생예측에 관한 연구가 많이 이루어졌지만, 노화 관련 질병 (전립선질환, 골다공증, 관절염, 백내장, 치매, 신경퇴행성 질환 등)의 질병 예측을 통하여 고령화 시대에서의 노화 억제 및 건강증진의 가이드도 필요하다. 기 개발된 여러 공공의 자료의 포괄적 활용과 함께 부족한 부분은 계속 개발해 나아가야 한다. 건강 증진에 맞추어 맞춤형 자료가 제공되어야 하며 질병 관련 자료도 맞추어 제공되어야 한다. 예를 들어 특정 암환자가 국립암센터에서 진료를 받은 경우 질병관리본부, 건강보험공단, 심평원 자료를 요약하여 critical 한 정보만 시급성·중요도·타당도에 의해 제공할 수 있을 것이다. 방대한 data를 제공하는 것 보다는 고급화된 정보를 요약해서 제공하는 것이 필요하다.



참고문헌





## 참고문헌

1. Aesun Shin et al, Risk prediction model for colorectal cancer: National health insurance corporation study, Korea. PLoS One. 2014; 9(2): e88079.
2. Bae CY, Kang YG, Suh YS, Han JH, Kim SS, Shim KW. A model for estimating body shape biological age based on clinical parameters associated with body composition. Clin Interv Aging. 2013;8:11-18.
3. Bae CY, Kang YG, Piao MH, et al. Models for estimating the biological age of five organs using clinical biomarkers that are commonly measured in clinical practice settings. Maturitas. 2013;75(3):253-260.
4. Bae CY, Kang YG, Kim S, et al. Development of models for predicting biological age (BA) with physical, biochemical, and hormonal parameters. Arch Gerontol Geriatr. 2008;47(2):253-265.
5. Beasley RP. Hepatitis B-virus: the major etiology of hepatocellular carcinoma. Cancer 1988;61:1942-1956.
6. Borkan GA, Norris AH. Assessment of biological age using a profile of physical parameters. J Gerontol. 1980;35(2):177-184.
7. Di Bisceglie AM. Hepatitis C and hepatocellular carcinoma. Semin Liver Dis. 1995;15:64-69.
8. Fleming et al. Risk prediction of hepatocellular carcinoma in patients with cirrhosis: The ADDRESS-HCC risk model. Cancer. 2014;120(22): 3485-3493.
9. Ford, Earl S., Wayne H. Giles, and William H. Dietz. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. Jama 287.3 (2002): 356-359.
10. Gene K. Ma and Uri Ladabaum, Personalizing colorectal cancer screening: A systematic review of models to predict risk of colorectal neoplasia. Clinical Gastroenterology and Hepatology. 2014; 12:1624-1634.
11. Giorgio Bedogni et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterology. 2006, 6:33.

12. Haydon GH, Hayes PC. Screening for hepatocellular carcinoma[Review]. *Eur J Gastroenterol & Hepatol*. 1996;8:856-860.
13. Hsu et al. Development of a scoring system to predict hepatocellular carcinoma in Asians on antivirals for chronic hepatitis B. *J. Hepatol*. 2018 Aug;69(2) 278-285.
14. Huang et al. Validation of the Fatty Liver Index for Nonalcoholic Fatty Liver Disease in Middle-Aged and Elderly Chinese. *Medicine*. October 2015, Volume 94, Issue 40, p e1682.
15. Ingram DK, Nakamura E, Smucny D, Roth GS, Lane MA. Strategy for identifying biomarkers of aging in long-lived species. *Exp Gerontol*. 2001;36(7):1025-1034.
16. Kohro T, Furui Y, Mitsutake N, et al. The Japanese national health screening and intervention program aimed at preventing worsening of the metabolic syndrome. *Int Heart J*. 2008;49(2):193-203.
17. Lau WY, Leow CK, Li AK. Hepatocellular carcinoma [Review]. *Br J Hosp Med*. 1997;57:101-104.
18. Markku Laakso, et al. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *Journal of Lipid Research*. 2017;58(1):481-493.
19. National Education Cholesterol Program Adult Treatment Panel III(NCEP ATP III). 2001.
20. Park SY, Yang YJ, Kim Y. Effects of nutrition education using a ubiquitous healthcare (u-Health) service on metabolic syndrome in male workers. *Korean J Nutr*. 2011;44(3):231-242.
21. Philipp Mitteroecker. Multivariate Analysis of Genotype-Phenotype Association. *Genetics*. 2016;202(4):1345-1363.
22. Suraj et al. Toronto HCC risk index: A validated scoring system to predict 10-year risk of HCC in patients with cirrhosis. *Journal of Hepatology*. 2018 vol.68 92-99.
23. Zoli M, Magalotti D, Bianchi G, Gueli C, Marchesini G, Pisi E. Efficacy of a surveillance program for early detection of hepatocellular carcinoma. *Cancer*. 1996;78:977-985.
24. 국가암등록통계 2016 <https://ncc.re.kr/cancerStatsList.ncc?sea>

25. 국가암검진사업 [https://www.ncc.re.kr/main.ncc?uri=manage01\\_4](https://www.ncc.re.kr/main.ncc?uri=manage01_4)
26. 권진원, 신상진 외. 국가간암암시검사가 간암 조기발견, 사망, 의료비용에 미치는 영향. 한국보건의료연구원. 2016
27. 김민호, 서주현, 이제영. 이상지질혈증을 예측하는 노모그램 구축 방법 연구, Journal of the Korean Data and Information Science society. 2019,30(5), 1063-1075.
28. 문창모, 박은정 외. 국가 대장암검진 추가검사 순응도의 요인분석과 임상적 성과 연구. 한국보건의료연구원. 2017
29. 안상훈, 한광협, 윤영훈, 홍성필, 백용한, 전재윤, 등. 한국인에서 간암발생의 위험요인 분석. 대한내과학회지. 2000;60:123-130.
30. 유종하, 육태미, 이상선, 계인선. 대장암 선별검사에 대한 근거생성 연구. 국민건강보험 일산병원 연구보고서. 2014
31. 임열리 외. ATP III의 진단기준에 따른 대사증후군의 유병률과 관련위험요인 분석-일개 대학병원 강증진센터 수진자를 대상으로. 가정의학회지. 2003; 24:135-143
32. 정재연, 한광협, 김동기, 안상훈, 송기준, 백용한, 등. 한국인에서 간세포암종 발생의 위험요인 및 개인별 간세포암종발생 예측모형. 대한간학회지. 2001;7:449-458.
33. 최윤정, 강중구, 이상현, 조용석, 육태미. 국가건강검진 자료를 이용한 대장암 조기검진 compliance 분석. 국민건강보험 일산병원 연구보고서. 2015



# 부 록



## 제 1차 포럼: 개인건강기록 경험과 전망

발표자: 서울아산병원 이재호 교수

### 1. 개인건강기록(PHR) 개요

개인건강기록(Personal Health Records, PHR)은 개인이 개인건강정보의 주체로써 건강정보의 저장, 활용, 접근권한을 관리하는 시스템으로, 개인 자신의 건강정보와 이에 대한 접근을 통제하는 것 이외에는 EHR과 유사하다. 다양한 의료기관으로부터 제공되는 개인의 진료정보와 개인 스스로 기록한 건강기록을 통합적이고 포괄적인 관점에서 바라본 개인의 평생건강기록이라고 할 수 있다. PHR은 기존 건강관리체계의 한계, 소비자 중심의 패러다임 변화, 보건의료서비스 정책 변화, 그리고 의료정보기술의 발전으로 등장하게 되었다.

PHR은 크게 독립형, 기관소속형, 연결형의 세 가지로 분류할 수 있다. 독립형은 개인이 직접 건강 정보를 입력하는 것으로 개인이 건강정보를 통제하고 쉽게 개발 및 적용하는 장점이 있으나 개인이 건강정보를 직접 입력해야하고 정보의 신뢰성이 낮으며 의료기관과 연계가 되지 않는다는 단점이 있다. 기관소속형은 개별 의료기관이 제공하는 개인 건강기록으로 현재 가장 많이 사용되는 유형이다. 의료기관 서비스와 연계성이 높고 주로 치료받은 의료기관과 연계가 가능하다는 장점이 있으나 개인이 건강정보에 대한 통제권이 낮고 의료기관마다 다른 PHR을 사용한다는 단점이 있다. 마지막으로 연결형은 다양한 의료기관에서 건강정보를 제공하고 개인이 관리하고 연계하는 것으로 건강정보의 통합이 가능하고 정보 전달 및 연계성이 높지만, 개발과 운영이 어렵고 이해관계자들이 많아 실제 구현이 잘 안되는 단점이 있어 최근에는 국가 차원에서 개발하고 있는 유형이다.

PHR은 환자와 가족들에게 건강정보의 통제권을 강화하고 건강관리의 주도권을 제공할 수 있으며 환자와 의료진과의 관계를 향상시킬 수 있다. 또한 환자안전과 보건의료서비스의 질을 향상시킬 수 있고 효율성과 편의성을 향상시킬 수 있으며 환자의 프라이버시를 보호할 수 있고 보건의료 비용을 감소시킬 수 있다.



## 2. 국내외 PHR 현황

미국에서는 2020년에 75%의 성인이 PHR을 사용할 것으로 예상된다. iBlueButton은 2010년부터 제공된 PHR로 가족의 건강관리가 가능하고 의료진에게 data를 제공할 수 있다. Apple Health Record는 EMR을 개발해서 협약된 병원에 제공하고 있다. Epic MyChart는 미국 EMR 시장의 70%를 차지하며, 환자 정보를 제공하고 병원에 신청하여 PHR을 이용할 수 있으며 patient portal을 제공한다.

OECD 회원국의 PHR/Patient portal 현황은 2016년 기준 80%의 국가가(15개국 중 12개국) 국민이 자신의 건강데이터에 접근할 수 있는 시스템을 구축하고 있다.(우리나라는 data가 없다) 환자포털연구에서 호주, 에스토니아의 월별 사용률은 1%이고 덴마크는 5%이며, 스웨덴은 인구의 38%가 portal 계정을 생성하였고 핀란드는 인구의 53%가 portal 접근이 가능하다. 만성질환자의 활용면에 있어서는 네덜란드의 경우 4%의 만성질환자만이 PHR을 사용하고 있고, PHR로 도움을 받을 수 있는 취약계층의 활용 문제에 대해서도 논의가 필요하다.

호주에는 My Health Record라는 PHR이 있으며, 에스토니아는 건강보험에 가입한 모든 국민에게 ID를 제공하여 진단검사 결과 및 해석, 투약 정보 등을 제공하고 있다. 캐나다에는 MyHealthNS라는 PHR이 있었으나 payment model 문제로 서비스가 종료되었다.

국내에는 2019년 서울대학교병원에서 개발한 PHR이 있으며, 그 외에도 건강iN, 에필케어 등의 PHR이 있다. 그 중 분당서울대병원의 PHR은 PHR과 EMR을 통합하여 개발하였으며, 마이헬스노트는 보험회사와 연계하여 당뇨관리서비스를 제공한다.

## 3. PHR 개발 및 운영 경험

내손안의 차트(MCMH1.0)는 응급환자용 PHR로 처음으로 real-time data를 제공하였다. 내손안의 차트 2.0은 만성질환, 건강관리를 통해 환자생성 건강데이터를 임상으로 활용하였다. 내손안의 차트 2.0의 접속자는 꾸준히 증가하고 있으며, 이용자들은 '나의 차트' 중 '검사결과'를 가장 많이 이용하는 것으로 분석되었다. 나의 차트와 건강관리 기능에서 지속 사용성을 보였으며, 젊은 여성(50세 이하)에서 가장 많이 사용하는 것으로 나타났고 그 다음으로 노인남성(50세 이상), 젊은 남성, 노인 여성 순이었다. 특히 PHR/PGHD의 사용 효과로 HbA1C를 제공하는 경우 지속적으로 사용하는 사람에서 유의미하게 수치가 감소하는 것으로 나타났다.

생활속의 응급 앱은 응급질환 정보, 혼잡도, 응급실 진료 흐름, 퇴실 안내문 등을

통해 환자중심적인 자료를 제공하고 있고, 응급실 진료 현장에서의 환자 교육을 통해 적시성을, 병원정보시스템 연계와 자동기록을 제공하여 효율성을 보여주고 있다.

이처럼 응급진료에서의 모바일 PHR을 통하여 응급실 방문 이전에는 응급실 방문 예약, 119 의사소통, 의료지도, 원격자문, 응급실에서는 응급실 대시보드, IoT 활용 모니터링, 응급실 진료정보 제공, 응급실 방문 이후에는 응급실 퇴실 요약, 퇴실 교육, 전원 알림, 외래 예약 등 공중보건 감시에 활용할 수 있는 가능성이 있으며, 환자 매개 응급진료정보교류에도 활용할 수 있을 것으로 생각된다.

응급상황에서 나의 건강지갑은 국내 최초 응급 PHR 서비스 사례로, 환자의 개인건강기록(PHR) 관리에 MyData 개념을 도입하여 환자 중심의 투명하고 보안이 보장되는 플랫폼을 구축하고, 이를 통해 의료정보의 비대칭성을 해결한 국내 최초 서비스 사례이다. 환자 개인의 응급실 기록(진료내역, 검사기록, 처방기록) 중 병원 간 공유가 필요한 핵심적인 데이터를 선별하고 이를 3개 병원이 공유할 수 있도록 표준화 적용하였으며, 해당 기록이 향후에 타병원 응급실, 외래진료로 활용될 수 있도록 확장성을 고려하였다. 또한 내려 받기 가능한 70여개의 병원 정보 항목 중 원하는 정보만을 선별적으로 내려 받기 하여 공유 가능하도록 하는 선별적 공유 서비스 모델을 구현하였다.

#### 4. PHR 전망과 방향

PHR의 최종 정착지는 한 개인의 평생 건강기록 서비스를 제공하고, 태아~사망, 일상생활~병원을 연결하며 소아/청소년, 젊은 여성에 대한 접근을 용이하게 하는데 있다. PHR의 지속 사용성을 위해 환자의 관점과 가치를 이해하고 임상 의사와 PHR service로부터 피드백을 받으며 의사결정을 지원하도록 해야 한다.

환자중심의 건강관리 도구로 활용될 수 있도록 환자참여와 의사소통의 도구가 되어야 할 것이며, 압, 당뇨, 만성 신질환과 같은 특정 질병 PHR로 활용하고 환자 자가 보고 결과와 환자매개 진료정보교류를 위해 활용되어야 할 것이다. PHR 인터페이스를 위해 전자의무기록, wearables, 공공보건 서비스, 디지털 치료제, 임상시험 등을 활용할 수 있을 것이다. 임상진료와의 통합 및 건강증진 서비스를 통해 지불/비즈니스 모델을 구축하고, 119 정보를 연계하여 응급상황에서의 PHR 서비스에 활용하고 공중보건 감시를 함으로써 PHR을 공중보건에 활용할 수 있을 것이다. 정확성, 저장, 사용, 통합 및 교류 등을 통해 안전하고 신뢰 있는 PHR platform을 구축할 수 있을 것이다.

## 제1차 포럼 토론자: 성균관대학교 생명물리학과 안선주 교수

이재호 교수님의 발표 내용은 PHR에 대한 훌륭한 evidence를 제공하였으며, 연령별, 질환별로 건강정보에 대한 어떠한 needs가 있는지에 대해 알 수 있었다. 전 세계적으로 PHR은 정책적으로 중요하며, real time으로 system에 접속하는 것이 목적으로 호주의 경우 모든 국민이 접속 가능하도록 하였다.

우리나라는 건강보험 data와 심평원 자료가 있으나 사회적·정치적인 이유로 편익을 제공하지 않고 있어 이를 어떻게 해결할지에 대한 토론이 필요하다. 이에 대한 개선 의지를 볼 수 있는 것으로 my health way는 digital health 분과에서 발표한 것으로 검진데이터, 약국 등의 정보를 활용하는 것에 대해 일정 부분 논의 중에 있으며, 진료 정보 교류가 real time으로 교류되어야 한다는 공감대가 형성되어 있고, 언택 소사이어티를 위해 원격진료 활성화에 대해 정부가 의지를 가지고 있다는 점이다. 국민건강보험공단의 건강in은 평균적·보편적·대표적 PHR로써 기능하고 기업은 지속적·연계적인 PHR service를 제공하는 등 역할에 있어서 차이가 있을 것으로 생각된다.

공단 data가 국민에게 건강증진, 건강개선 서비스를 제공할 수 있을 것인가에 대해서는 목표를 세분화해서 연령, 지역, 질병, risk factor 별로 개발이 필요할 것이며, 현재 시점에서 할 수 있는 서비스가 무엇인지 define 하는 것이 필요하다. 예를 들어 국가 검진을 통해 한국인 평균 data와 개인의 data를 비교하여 보여줌으로써 어떤 건강행동을 해야 다른 disease를 줄일 수 있을 지에 대한 정보를 제공하여 개인에게 특화된 건강정보 서비스를 제공할 수 있을 것이다.

어떤 data가 필요할 것인가에 대해서는, 예를 들어 특정 암환자가 국립암센터에서 진료를 받은 경우 질병관리본부, 건강보험공단, 심평원 자료를 요약하여 critical 한 정보만 시급성·중요도·타당도에 의해 제공할 수 있을 것이다. 이처럼 방대한 data를 제공하는 것 보다는 고급화된 정보를 요약해서 제공하는 것이 필요하다.

PHR에 정보를 입력함으로써 정보의 연계성을 갖고 정보를 환자에게 날것으로 제공하지 않고 요약해서 제공할 수 있을 것이다. 이를 통해 AI powered PHR, 즉 질문하면 답을 주는 PHR을 개발하여 단계별로 발전하여 양질의 서비스를 제공할 수 있다.

PHR 개발을 위해서는 캠페인이 필요한데, 건강in에서 정보공개 우수사례로 상을 받은 적이 있으나 이후 PHR의 필요성에 대한 의사결정에 혼돈이 있어 진척이 없는 상황이다. 시대가 변화했으므로 패러다임의 변화가 필요하다는 로드맵이 필요한 상황이다.

마지막으로 JKMA에 발표된 생활치료센터 사례를 보면, softnet이 개발한 PHR로 untacked monitoring(fever 시 방 번호의 체온부분이 빨강게 변하거나 대상자가 증상을 표시할 수 있도록 해놓음)을 통해 직접 환자를 보지 않고 증상 관리가 가능한 것으로 보고되어 PHR이 무궁무진하게 발전할 수 있는 가능성을 보여주었다.

## <질의 및 응답>

Q1. (안병윤) : 건강in은 앞으로 어떻게 PHR을 할지에 대해 논의중이며 국가 PHR과 EMR PHR을 나눠야 한다는 것에 공감한다. 1) 모바일 PHR에 왜 응급환자를 다뤘는지에 대해 궁금하고, 2) 작은 병원에 갔을 때는 어떻게 활용해서 건강관리를 할 것인지, 3) 각각의 PHR이 만들어지면 control은 어떻게 할 것인지에 대한 의견 부탁드립니다.

A1. (이재호) : 암, 당뇨, 심장질환 등 만성질환이 있으면 응급실에 자주 오게 된다. 이때 PHR을 활용하여 신원을 파악하고 환자 연계 등에 사용할 수 있고, 응급실 방문 전/응급실에서/응급실 방문 후에 기본적 건강정보에 더하여 특정 PHR 서비스가 필요하다. 작은 병원에서는 불가능한 상황이고, 국가차원에서 건강보험공단이나 심평원에서 정보를 제공하는 것에 대한 안전망과 검증이 필요하며, 보건의료계 또는 과학기술부 등 어디에서 control 할지에 대한 논의가 필요하다.

A2. (조경희) : 개인 주도형 건강데이터(PHR) 이용 활성화 BPR/ISP 컨설팅 사업에 6개월 동안 5억여원의 예산이 투자 될 예정이며, 개인 중심 의료데이터 통합 및 연계 체계를 구축하고 개인 중심 의료데이터 기반 서비스 개발을 지원하고 지속가능한 혁신을 위한 거버넌스 구축을 추진하고 있다. 또한 안심하고 자신의 의료데이터를 주도적으로 통합·활용하기 위해 국가적 아젠다로 지속적·체계적으로 추진할 계획이며 총 3단계로 사업이 진행될 예정이다.

## 제 2차 포럼: 주제 1) 데이터 관점의 전자의무기록

발표자: 인제의대 강성홍 교수

### 1. 전자의무기록을 왜 하는가

지역사회 돌봄을 할 때 의무기록을 위해 필요한 것이 Community Health Record(CHR)이다. 의무기록을 병원에서만 쓰는 것이 아니라 community로 나가게 되면서 의무기록의 2차 이용이 중요하게 되었다. 2017년 CDC에서 발표한 건강의 결정요인은 지금까지의 건강 요인이었던 임상진료(20%)와 건강행태(30%)에 환경(10%)과 사회/경제(40%)를 추가하였으며 이 요인을 해결하기 위해서는 정부의 역할이 필요하며 정부를 포함하는 것이 CHR이다. 미국의 경우 만성복합질환을 가지고 있는 5%의 환자에서 전체의 45%의 의료비를 지출하고 있으며 이 환자군을 어떻게 관리 할 것인가에 대한 연구로 CHR이 등장하게 되었다. 또한 지역사회 노인통합돌봄사업을 통해 시설에서 지역사회로 확장되면서 여러 지역, 여러 직종에서 CHR이 필요하게 되었다. 이처럼 community care 시대, 즉 병원 밖에서의 관리를 위해서는 CHR이 필요하다. CHR의 3대 요소는 건강관리, 지역사회, 공중보건이며 개인의 건강관리와 지역사회의 data가 모여서 공중보건으로 문제를 해결해야 한다.

CHR은 자격을 갖춘 의료 서비스 제공자를 위한 환자의 의무기록으로서, 지역사회의 여러 의료기관에서 수집된 환자의 통합된 의무기록을 말한다. 즉, 건강에 영향을 미치는 사회적, 신체적 및 생활양식에 대한 개인 및 지역(집단)에 관한 건강정보로서 주로 보건 의료서비스 제공자를 위해서 구축된 지역 주민에 대한 건강기록이다. CHR의 구성요소로는 임상정보(EHR, 환자포털), 공공보건기관(설문조사, 감시), 사회서비스(사회복지서비스), 집단건강(돌봄관리), 비 임상 건강 관련(범죄, 대기오염), 보험회사(청구) 등이 있다. CHR의 데이터 항목에는 개인 단위의 데이터인 My Health Bank와 지역단위의 지표 데이터인 건강 및 질병지표가 있다.

### 2. 무엇이 가장 큰 문제인가

EHR data의 문제점은 EMR data의 80%가 코드화 되어 있지 않은 경과일지, 영상 기록지 등과 같은 비정형 data라는 것이다. 비정형 data를 어떻게 coding 할 것인가가 중요한 issue 이다. 또한 정형 data도 품질이 좋지 않다. 병원의 20%의 정형 data

중에서도 청구 목적을 위한 10%정도의 정보만 이용 가능하여 이 정도의 정보만으로 신약 개발이 가능 할 것인가 하는 문제가 있다. 그리고 환자 개인정보보호는 중요한 문제인데 개인의 프라이버시를 중요시 하는 것과 환자 안전 및 보건의료 수행을 중요시 하는 것 사이에서 적정 수준을 유지해서 데이터를 관리하자는 것이 데이터 거버넌스의 핵심 개념이다.

### 3. 어떻게 해결할 것인가

기술적인 문제도 있지만 정책적인 문제가 있다. 이를 해결하기 위해 데이터 관리 정책의 수립→데이터 관리 전략 수립→데이터 거버넌스 기반의 데이터 관리가 필요하다. OECD의 EMR 데이터에 대한 접근은 다음과 같다. 먼저 2012년에는 EMR의 2차 활용이 점점 중요해지는 시대로 OECD에서는 EMR 데이터를 공공재로 보고 EMR 데이터를 이용하여 의료의 질향상을 도모하기로 하였다. 또한 2013년에는 의료의 질에 대한 지표는 환자 코호트 데이터에 기반한다는 전제로 통합된 EMR 데이터를 의료기관 간에 진료 정보를 공유하고자 하였고, 진료정보 공유를 위해서 표준, 데이터 연계, 개인 정보 보호가 매우 중요해지므로 보건의료 데이터 거버넌스 구축이 필요하게 되었다. 2015년에는 환자의 동의를 받지 않은 EMR 데이터를 어떻게 활용할 것인지에 대해 retrospect 코호트로 EMR 활용방안을 검토하였고 보건의료 데이터 거버넌스 관점에서 각국의 현황을 파악하여 보건의료 데이터를 사용함으로써 얻을 수 있는 이익은 극대화하고 위험은 최소화 하자는 보건의료 데이터 거버넌스를 구축하게 되었다. 2017년에는 Health data governance를 발표하여 가이드라인 원칙을 제시하였다. OECD Health Data Governance에서는 참여 및 참가, 기관들 간 협력, 정보시스템의 성능 검토, 개인에 대한 명확한 정보제공, 사전 동의(informed consent) 및 적절한 대안, 건강정보 사용에 대한 적절한 검토 및 승인 절차, 공공정보 체계의 투명성, 프라이버시 보호와 정보 보안을 보장, 감독과 평가체계, 교육훈련 및 기술개발 방안 수립, 통제 및 안전장치의 구현, 개인 보건 자료를 처리하는 기관 등에 관한 지침을 제시하였다. 2019년에는 OECD의 EMR 데이터에 대한 접근의 마지막은 필요한 시점에 필요한 사람에게 필요한 의료정보를 제공해주자라는 핵심 agenda를 가지고 인센티브를 동반한 의료정보 정책, 보건의료 데이터 거버넌스, 정보화된 조직 인력 양성에 대한 내용을 제시하였다.

미국에서는 데이터 관리정책을 위해 2018년 3월에 President's Management agenda를 준비하였고 2019년 말에 Action이 만들어 졌으며 2020년 3월에 COVID-19에 대응하

도록 Action plan을 수정하여 발표하였다. 추진배경은 정부기관이 보유한 공공데이터를 더 많이 개방 및 공유하고 데이터 프라이버시 문제를 최소화 하여 이를 민간, 연구자, 기업이 2차 활용할 수 있도록 범정부적인 노력이 필요하다는 것이었고 이를 위해 모든 기관의 데이터 표준화, 상호 운영성, 데이터 역량 개발을 위한 중장기 범정부차원의 데이터 전략인 연방데이터 전략의 원활한 이행을 위한 세부 이행 과제를 수립하였다. 10개의 원칙과 40개의 과제, 20개의 수행단계를 발표하였으며 2020년 주요 세부 추진 과제는 다음과 같다. 기존 과제 수행 시 필요한 데이터의 선제적 발굴, 다양한 거버넌스 기구 구성, 데이터 및 데이터 관련 인프라 성숙도 평가 진행 등이 있으며, 성숙도 평가 지표는 ‘기관 최고 데이터 책임자와 다른 이들은 연방에서 제공하는 교육 또는 기술지원에 참여하였다’였다. 이 외에도 AI 연구 및 활성화를 위한 데이터 개선에 대한 과제로, AI도 정형 data가 있어야 가능하므로 20%의 정형 데이터로 가능하기 위해서는 AI로 돌릴 수 있는 data set을 만들어야 할 것이다. 그 외 세부 추진 과제로 데이터 윤리 프레임워크 개발, 데이터 품질 관리가 중요하므로 데이터 품질 측정 및 보고 가이드 개발 등이 있다.

미국과 한국의 데이터 관리 전략에는 차이점이 있다. 미국은 좋은 데이터 셋을 만드는 데 목표를 두는 반면 한국은 데이터 개방 및 활용이 목표이다. 미국은 데이터 거버넌스인 반면 한국은 정보 거버넌스이며, 미국은 관리자를 Chief Data Officer(CDO)로 지칭하는 반면 한국은 Chief Information Officer로 부른다. 사업 내용에 있어서 미국은 데이터의 질적 수준 향상을 중시하고 한국은 개방 진수를 중시한다는 차이점이 있다.

ERP(Enterprise Resource Planning)란 계획수립, 제조, 영업, 마케팅, 재무, 인사관리 등을 포함한 기업의 제반 요소를 통합하고 기업 활동을 위해 사용되는 기업 내의 모든 인적, 물적 자원을 효율적으로 관리하여 궁극적으로 기업의 경쟁력을 강화시켜주는 역할을 하는 통합정보시스템으로, 정보를 활용하여 조직의 자원을 효율적으로 관리하는 방법이다. EIM(Enterprise Information Management)은 정보를 주요 자산으로 보고 정보를 효율적으로 관리하는 방법론으로 정보를 잘 관리하게 하고 더 나아가 양질의 정보로 ERP도 효율적으로 이루어지게 한다.

OECD에서는 data governance와 information governance를 포함하는 개념을 사용하는데, 이에 대한 엄격한 구분이 필요하다. data governance란 data를 잘 관리하는 것이고, information data는 들어가 있는 data를 잘 활용하는 것이다. 거버넌스란 정부와 시민사회가 협력하여 사회 문제를 해결하는 것을 말한다(하나로 가지 말자는 개념).

데이터 거버넌스란 데이터 관리에 있어서 정부 주도의 통제와 관리가 아니라 데이터 관리 전문가들이 스스로 통제하고 관리하는 것으로, 표준과의 차이점은 국가의 데이터 표준은 하나여야 하지만 데이터 거버넌스는 여러 개 있을 수 있다는 것이다. 보건의료데이터 거버넌스는 의료기관이 자발적으로 환자의 의무기록/의료정보 데이터와 의료기관의 운영에 관련된 데이터에 관한 정책과 프로세스, 프라이버시 방안, 비밀보장 방안, 데이터 질 관리 방안 등을 만들고 준수하는 것이라 할 수 있다.

보건의료 데이터 거버넌스 방법론에는, 예를 들어 KPMG는(컨설팅 회사) 가치 기반의 데이터를 만들기 위해서 데이터 거버넌스 방법론을 사용하고 있다. 정보관리와 거버넌스 전략을 위해 인력(조직구조, 역할과 책임, 문화와 의사소통), 프로세스(데이터 자산관리, 프로세스 활성화, 정의와 표준), 기술(데이터 질 관리, 측정항목과 모니터링, 도구와 기술)을 통해 데이터 거버넌스 정책과 도구를 개발 할 수 있다. 각각에는 여러 가지 요소가 있으며 level 1~level 5까지 구성되어 있다. 이를 한국 현실에 맞게 접근하는 것이 중요하다. 우리나라에서 채택한 방법론에 의한 보건의료 데이터 거버넌스 영역에는 데이터 생애주기 관리, 데이터 아키텍처, 메타 데이터 관리, 마스터 데이터 관리, 내용관리, 데이터 보안 관리, 비즈니스 인텔리전스, 데이터 질환리, 용어관리 등이 있고 차이점은 콘텐츠를 중요시하기 때문에 ‘내용’이 들어갔다는 것이다. 즉, 사람-프로세스-내용-기술을 통해 기획-조직-통제하게 된다. 보건의료 데이터를 효율적으로 활용하기 위해서는 표준만을 해서는 안 되고 보건의료 데이터 거버넌스를 해야 한다. IBM에서 자연어 처리하는 프로그램을 개발하였고, 국내에서도 개발 중이다. 데이터 질 관리는 의료정보망에서 미비기록 관리를 통해 충실히 기록하도록 유도한다. 비즈니스 인텔리전스에서는 데이터의 시각화를 전략으로 삼았다. 환자 데이터는 밑으로 가는 히스토릭한 데이터로 밑으로 데이터를 붙이는 process mining을 이용한다. Dynamic capacity management는 시시각각 변하는 상황을 시간대별로 관리하는 것으로 이를 위해서 포항공대 산업공학과에서 시뮬레이션 tool을 개발하였다. GIS+Regression은 지역별로 회귀계수가 나오며, 이를 이용하여 지역별 보건사업에 대한 시뮬레이션 프로그램을 개발할 수 있다. 예를 들어 종속변수를 흡연율로 했을 때 공단의 지역사회 지표 데이터로 독립변수로 넣어 GW로 돌리면 회귀계수가 지역별로 다 다르게 결과가 나와서 지역별로 상대적으로 어떤 변수가 상대적으로 중요한지에 대한 결과를 얻을 수 있다. 내용관리에 있어서는 free text로 해놓으면 쓰기는 좋지만 내용관리가 되지 않기 때문에 check box form으로 항목을 만들어서 내용관리가 가능하도록 하였다.



#### 4. 언제 해결 할 것인가

미국은 '범정부 총력전'을 하고 있어 국제적 동향에 따라 우리나라도 논의가 필요하다.

#### 5. 어디서 해결할 것인가

거버넌스이므로 하나로 갈 수가 없다. 병원은 병원대로 지역사회는 지역사회대로 형편에 맞게 원칙을 정해서 해결해야 한다. 국가별로, 지자체별로, 병원별로 만들어서 하나씩 가야한다.

#### 6. 누가 할 것인가

모든 분야는 발전하면 세분화 된다. 보건의료정보의 영역은 HICT(기술적인 것), HIM, HI로 나뉜다. HIM(Health Information Management)은 보건의료정보의 체계적 수집 및 관리 하는 것으로 보건의료의 품질, 안전성, 효율성 향상에 필요한 보건의료정보의 체계적 수집, 분석 및 보호의 역할을 한다. HI(Health Information)은 보건의료정보의 과학적 분석 및 활용하는 것으로 연속적인 보건의료 전달에 적용되는 정보, 체계 및 기술을 말한다. HICT(Health Information and Communication Technology)는 보건의료정보관련 정보통신기술지원으로 보건의료정보 및 시스템 관리의 기술적 측면에서 하드웨어, 소프트웨어 기술 지원의 개발 및 관리를 담당하고 인프라, 기술 및 도구를 사용하여 디지털 형식으로 정보를 교환하는 보건의료정보 개발, 관리, 보안 관리를 한다. 이 세 영역을 통합하는 IG(Information Governance)는 상위 수준의 관리감독이 요구되는 전략적 자산으로서의 정보에 중점을 두고 책임성, 투명성, 무결성, 보호, 규정준수, 유용성, 보유, 처분을 원칙으로 한다.

**결론 :** 의료분야의 4차 산업혁명의 핵심은 보건의료 데이터 관리이고 국가 및 기관차원의 보건의료데이터 거버넌스를 구축해야하며, 데이터 분석결과를 보건의료서비스의 개발 및 관리에 이용해야한다. 또한 보건의료데이터를 이용하여 보건의료시스템을 강화해야하며 고용을 창출하는 보건의료 4차 산업혁명이 필요하고 CDO(Chief Data Officer)가 양성되어야 한다.

## <질의 및 응답>

Q1. (조경희) : 의원에서는 의원 수준에 맞춰서, 병원급은 병원 수준에 맞춰서 의료 정보를 이용하게 되는데, 데이터 거버넌스를 어떻게 풀어 갈 것인가. 지역 사회에서 방문 간호사가 데이터를 이용하기 위해서 CDO가 필요한데 이에 맞는 데이터는 어떻게 할 것인가.

A1: (백설경) : 현재 병원 데이터를 가지고 이용하고 있는데 CHR이 나오면서 지역사회에서 필요한 data도 만들어야 하고 더 나아가서 국가적으로 데이터 거버넌스를 해야 하는 것도 있다. 지금 있는 것은 병원 데이터만 있기 때문에 가정간호는 병원보다는 지역사회에서 필요하다. 또한 지역사회에서 필요한 부분은 환경 부분도 포함한다. 이러한 것들이 축적되기 위해서는 병원에서는 data가 축적이 되어있고 표준화 되어 있지만 CHR로 봤을 때는 병원 data를 흡수해서 넓히고 국가에서 병원 data와 CHR data와 그 외 data들을 모아서 전체적인 data governance가 마련 될 수 있을 것이라고 생각한다.

Q2. (조경희) : 공단 big data는 청구 data와 검진 data들이 모여져 있다. 이런 것들이 오늘 얘기한 주제와 연계가 될 것이라고 생각하는가.

A2. (백설경) : 공단 data가 가장 중요할 수 있다. 병원에 없는 data들이 모여져 있기 때문에 정제된 data들이 정리가 된다면 충분히 활용할 수 있을 것이라고 생각한다.

## 제 2차 포럼: 주제 2) 진료정보교류 표준

발표자: 삼성병원 이병기 교수

방문 간호사가 환자의 정보를 한번만 조회해도 모든 정보를 볼 수 있기 위해서 가능한 방법은 캐나다나 호주에서처럼 National EHR을 개발하는 것이지만 이것은 아무도 원하지 않는다. 이렇게 하지 않고 정보를 수집하기 위해서는 제한된 영역에 있는 data만 표준화해서 그것을 주고받는 것이 HRE가 되는 것이고 실제 우리나라에서 진료정보교류 사업을 하고 있다.

## 1. Health IT interoperability standards - paradigms and structures

표준을 왜 하는가. 예를 들어서 여러 회사에서 만든 EMR system이 기능적으로 호환 가능하도록 상호 운영성을 갖도록 하기 위해서이다. interoperability classification에는 semantic interoperability, Technical interoperability(data를 주고받을 수 있는지, 기술적인 것), Functional interoperability(기능적, process적인 것)가 있다.

진료정보교류사업에서 임상서식을 주고받는 것(use case)을 위해 정보들이 활용되기 위해서는 구조화된 data(codified data), 숫자나 이름을 어떻게 표현하고 나열하고 grouping 할 것인지가 Information Model이다. 즉 use case independent 하게 이렇게도 저렇게도 쓰는 것이 information model 이다. Refined Information model은 목적, 용도에 맞게 specific 하게 refine 할 때 사용하는 정보 model 이다.

문서표준, 메시지 표준은 특정기관에서만 쓰는 것이 아니라 국제표준이 되어 모든 나라가 쓸 수 있어야 한다. 이렇게 되면 specific은 빠지게 된다. 각 나라에서 쓰는 진단 코드가 다르고 이름을 표기하는 방법이 다르기 때문이다. base 교환 표준(국제문서규격 표준)에 국가 레벨의 requirement가 추가되어 National level의 구현 가이드가 만들어지고 use-case level로 각 서식에 대해 구현가이드가 있어 실제로 쓸 수 있도록 profile을 하게 된다. 국제표준을 그대로 가져다 쓰는 것이 아니라 실제로 쓸 수 있도록 하는 profiling 하는 process가 추가되게 된다.

Health Interoperability Paradigm은 정보가 어떤 방식으로 교환되는지에 대한 것이다. Message는 사람이 전혀 개입하지 않고 application 혹은 system 간 정보 교환이 목적이다. 예를 들어 검사 처방을 하게 되면 검사실 정보시스템에 요청이 가게 되고 검사 결과가 나오면 자동으로 요청한 시스템으로 결과를 보내준다. 정보가 오갈 때 사람이 개입이 되지 않고 컴퓨터가 해석을 하게 된다. Document는 사람의 정보 consumption을 위해 만들어졌으며 사람을 대상으로 한 정보 교환이 목적이다. 정보는 사람의 가독성이 보장되고 진료정보센터의 4가지 서식이 이에 해당한다. API는 서버와 application 간 대화형(interactive) 정보 교환하는 것이다. 양쪽에 사람이 있을 수도 있고 AI처럼 기계가 서비스를 제공할 수 있으며 interaction이 발생한다. 최근에 스마트 기기가 등장하면서 API가 중요해졌다.

## 2. CDA Intro

HL7은 의료정보표준을 개발하는 비영리기관으로 더 이상 새로운 버전은 개발하지 않는다. CDA는 임상문서에 관련된 것으로 문서 표준이다. FHIR는 HL7에서 active하게

개발하고 있는 것으로 계속 version up 되고 있다.

HL7 CDA(Clinical Document Architecture)는 임상문서 교환을 위해 문서의 구조를 정의한 표준으로 RIM을 기반으로 하고, XML로 작성한다. 다양한 교환서식을 정의할 수 있는 서식을 제공하는 것이 CDA이며, 사보원 진료정보교류사업에서 사용하는 base 표준이다. CDA templates에는 CCD(continuity of Care Document)와 C-CDA(Consolidated CDA)가 있다. CCD는 서식의 모임으로 HL7 V3 CDA 포맷에 CCR(Continuity of Care Record)의 장점인 기능적인 부분을 통합하였고, C-CDA는 여러 서식을 한꺼번에 모아서 개발하고 HL7, IHE, HITSP에서 만들어진 CDA template들을 통합하기 위한 것으로 CDA template 라이브러리를 포함하고 있다.

CDA는 Header와 Body로 이루어져있다. Header에는 Meta data가 들어가며, 환자의 정보, 어떤 의사가 기록했고, 어느 의료기관에서 언제 했는지가 이에 해당한다. Body에 실제 임상정보가 기록된다. 이는 사람을 위한 것이므로 Human-readable은 반드시 필요하지만 Machine-readable도 optional하게 가능하다. human-readable 정보는 free text 형태로 data가 들어가 있어서 사람이 눈으로 확인은 가능하지만 컴퓨터가 분석하는 것은 불가능하다.

CDA의 level 1은 free text로만 이루어진 것이고, level 2는 개별 data는 free text이지만 grouping된 section의 의미를 code를 이용해서 표현한다. level 3는 detail한 level까지 code로 나타낸다. 현재 나오는 구현 가이드는 level 3까지를 지향하고, 이를 뒷받침하기 위해서는 EMR에서 generate가 되어야 하고 EMR이 보조가 되어야 한다. 이를 위해서 보건복지부에서 EMR 인증 기준을 높이는 과정 중이다. Consolidated CDA는 미국에서 쓰고 있으며 continuity of Care Document(CCD), Consultation Note, Diagnostic Imaging Report, Discharge summary, History of Physical Note, Operative Note, Procedure note, Progress note와 같은 8개의 서식을 define하고 있는 것을 종합적으로 말한다.

진료정보교류 사업을 하면서 4종 서식 구현가이드의 고도화에 대해 작년에 지침을 개정하였다. 기존 지침서를 따라서 구현을 하다 보면 CDA가 아닌 case가 나오게 되어 이를 바로 잡기 위해 개정하였다. 또한 미국에서 사용하는 구조화된 교류 서식이 gap이 있어 이를 구조화 및 고도화 하는 방향으로 만든 것이 K-CDA이다. 4종 서식(의뢰서, 회송서, 요약지, 임상의학정보)은 바꾸지 않고 고도화 하였다.

### 3. FHIR Basics

FHIR(Fast Healthcare Interoperability Resources)에서 'Fast'는 개발이 쉽고 빨라진다는 의미(운용 성능이 좋아진다는 것이 아니라)이고 'Resource'는 웹브라우저 주소창에 URL(Universal Resource Locator)을 넣게 되는데 이처럼 인터넷 웹에서 모든 것은 resource 형태로 address화 한다. 이처럼 FHIR에서는 정보가 URL 형태로 표현이 된다. FHIR는 2010년부터 개발이 시작되었다. CDA는 더 이상 개발이 되지 않고 있으며, FHIR는 V2, V3, CDA가 하지 못하는 것도 가능하고 표준 적용이 빠르기 때문에 FHIR를 집중적으로 개발하고 있다. FHIR의 구성요소에는 2 main part로 content model과 exchange specification이 있다. content model(Resources)은 의무기록상의 환자 정보(성명, 주소 등)정보를 말하고 exchange model은 실제로 정보가 어떻게 교환되는지에 관한 정보이다. 일례로 CDA는 서식만 define 하고 주고받는지에 대한 정의가 없지만, FHIR는 어떻게 주고받는지에 대한 것도 같이 표준 안에 정의되어 있다. 이뿐만이 아니라 knowledge, Decision support, Quality measures, being used for persistence(EMR의 data를 native FHIR resource를 기반으로 해서 만들자) 등으로 extending 하고 있다.

기존의 information model은 모든 것을 다 표현하려고 하기 때문에 복잡하다. EMR의 80%는 쓰는 사람이 거의 없기 때문에 실제 use의 20%만 FHIR 표준으로 정의하고, 나머지는 extension 해서 쓰면 된다. 대신 모든 사람이 다 같이 많이 쓰는 20%를 표준화하게 되면 manageable하게 된다.

### 4. FHIR Projects

The Argonaut Project는 US national profile을 만들기 위한 것으로, 여러 종류의 Technology Vendor에서 참여하여 US core를 만드는데 참여하였다. SMART on FHIR에서 SMART는 Substitutable Medical Applications, Reusable Technologies로 EMR, EHR application을 만들면 다른 곳에서 다 쓸 수 있도록 하는 것이다. 예를 들어서 안드로이드는 기기는 다양하지만 app은 하나만 만들면 모든 곳에서 돌아간다. EMR은 다 다르기 때문에 현실적으로 하나의 PHR로 여러 병원에서 열어보는 것은 불가능하지만 FHIR 표준을 이용해서 데이터를 가져오게 되면 실현 가능할 수 있다.

Medical APP gallery에 등록을 하면 여러 종류의 EMR system은 FHIR을 가정하고 만들게 되고 FHIR에서 표준화 하고 vendor에서 연동을 가능하게 하면 gallery에서 여러 다른 EMR system을 열 수 있게 된다. 기존 EMR system을 기능적으로 extension 하는 것은 SMART on FHIR를 통해서 할 수 있다.

Decision support 관련하여 CDS Hooks라는 개념이 있는데 DUR을 연상하면 된다. 처방된 data가 심평원에 넘어가서 경영동기를 test 하는데, 경영동기 이외에 decision support를 제공하는 경우를 생각해보자. 예를 들어 drug allergy가 있는 경우 penicillin allergy가 있어도 brand name이 올라와 있기 때문에 약 정보에 대한 전문적인 knowledge base가 올라와 있어야 한다. 심평원 DUR 인터페이스, allergy 관련된 인터페이스 서비스를 add 할 때마다 다른 인터페이스를 구현해서 EMR에 넣어야 한다. 이런 것들을 다 아우를 수 있는 프레임 워크가 있는 것이 좋겠고 이것이 CDS Hooks 이다. 대부분 CDS service는 어떤 환자인지 context data(Hook context data)가 필요하고 미리 보내는 data(Prefetch FHIR data)가 필요하며, 서비스별로 case별로 원하는 data가 다를 경우(Retrieve FHIR data)의 data를 모두 FHIR 표준으로 표현하자. 이 프레임워크를 깔고 이 service에 맞춰서 구현하게 되면 어떤 EMR에서도 활용 가능할 수 있다.

이외 FHIR 관련된 project에는 Da Vinci(청구 관련), Blue button 2.0, CARIN(요양기관으로부터 받은 PHR data를 다시 의료기관으로 보내는 것), Codex(cancer data), Gravity(마약성 진통제 때문에 사회적 문제가 되어 social behavior 변화에 활용) 등이 있다.

## <질의 및 응답>

- Q1. (조경희 교수) 무엇에 대한 allergy가 있다에 대한 정보는 code 된 data가 있어야 활용 가능할 것인데, '사과'에 대한 allergy가 있다는 수준으로 code가 있는지?
- A1. Level 3 수준으로는 EMR 상 기록에 없을 수도 있고, free text 라서 codified가 안되었을 수도 있다. data의 source는 EMR이며 이 source data를 이용해서 exchange 하게 된다. 항상 EMR 인증 기준이 선행이 되어야 하고 exchange 표준은 이를 따라서 발전하게 된다. CDA-FHIR 관련해서 진료정보교류사업은 4종 서식 중 진료정보교류서식 중 실질적으로 안쓰는 서식은 요약지로, 이는 process에 없다. PHR 관점에서 보면 요약지가 제일 유용하다. FHIR은 기관 대 기관 보다는 환자 중심으로 뭔가를 하기 위해 좋다. 보건 의료정보원에서 올해 시행하는 2차 EMR 사업에서는 기존의 revise한 CRS(Level 3)를 FHIR 바꾸는 작업을 하고 있다. 이렇게 되면 정보는 CRS와 같은데 방식은 FHIR 방식으로 PHR을 이용할 수 있다.

# 제 3차 포럼: 주제 1) 진료정보교류사업 현황과 방향

발표자: 사회보장정보원 엄재선 단장

## 1. 진료정보교류사업 개요

### 가) 정의

진료정보교류사업이란 개인 정보 제공에 동의한 진료기록을 국민이 원하는 곳(의료기관)에 전자적인 방식(안전성·정확성)으로 전송하여 의사가 진료에 참고(신속성·연속성) 할 수 있도록 제공하는 서비스로 의료법 제 21조의 2(진료기록의 송부 등)에 법적근거가 있다.

### 나) 추진 배경

국민의 입장에서는 환자가 직접 진료를 복사/전달하는 불편을 해소하고 불필요한 검사·영상촬영 등을 최소화하여 환자의 의료 효용을 증가 하는 것이고, 의료기관은 환자의 진료 이력을 참조해 신속하고 빠르게 진료에 대응할 수 있고 협력 의료기관 간 협진 네트워크 강화로 의료서비스 품질을 향상시킬 수 있다는 것이다.

### 다)추진 경과

2006년부터 추진되었으며, 보건복지부와 분당서울대병원에서 연구 및 시범사업을 추진하였다. 2016년에 의료법이 개정되어 법적 근거가 마련되었고 2017년에 제도가 갖춰지고 본 사업을 할 수 있는 기반이 마련되었다. 2018, 2019년 진료정보교류사업 추진을 위한 거점의료기관과 문서저장소가 전국적으로 구축이 되어 네트워크가 완성되었다. 2020년에는 진료정보교류 서비스 연계를 확대하고 있다.

### 라) 추진 현황

진료정보의 안전한 보관을 위해 14개의 거점 병원의 문서저장소와 1개의 공공저장소를 구축하였다(제주도는 공공저장소 활용). 52개의 거점 의료기관이 있으며 거점 41개, 단위기관 11개의 의료기관이 있다. 참여 의료기관은 6,439개 병원이 참여하고 있다. 올해는 치과 18000여개 중 236개소가 참여 예정이다.

4년 동안 11.2만 명의 환자가 21.4만 건의 진료정보교류서비스를 이용하였고 매년 200% 이상 진료정보교류 실적 증가가 있었다. 이는 상급종합병원 중심의 협력체계, EMR 개발사의 적극적인 기술에 의한 것으로 생각된다.

영상정보교류는 2019년 본 사업 추진에도 불구하고 연간 약 3만 건의 영상

정보교류서비스가 이용되었다. 영상정보교류 시 가산수가(+4천원)가 신설되어 10월 5일부터 개시 될 예정으로 이용률 증가가 예상된다.

#### 마) 추진 성과

전국 단위 진료정보교류 네트워크를 완성하여 기반을 구축하였고, 의료기관 지원 체계 및 인센티브 제도를 마련하였으며, 의뢰·회송 가산 수가를 신설하고 EMR 인증 제를 통해 공공재로써의 진료정보교류 영역을 확대하여 심평원·중앙응급의료센터·근로복지공단 등 정보를 연계하여 교류를 활성화 하였다.

### 2. 진료정보교류 주요 서비스

#### 가) 진료 의뢰

환자의 진단·치료를 목적으로 상급 의료기관으로 의뢰할 수 있으며, 병원 방문을 예약하고 개인정보제공에 동의하면 진료의뢰서, 진료기록(진단, 검사, 약물, 수술, 알리지 및 영상 등) 등을 제공한다.

#### 나) 진료 회송

환자의 상태가 호전되어 의뢰 받은 의료기관 또는 타 의료기관으로 전원 시 진료회송서, 진료기록(진단, 검사, 약물, 수술, 알리지 및 영상 등) 등을 제공한다.

#### 다) 진료기록 요약지

의료기관에 내원한 환자의 진료기록을 요약지 형태로 작성하여 문서 저장소에 보관하여 제공한다.

#### 라) 영상 정보 교류

CT, MRI 등 영상정보파일과 판독 소견서를 제공한다.

### 참고) 진료정보교류 개념

거점 문서 저장소는 물류 창고 개념이며 진료기록전송지원시스템은 headquarter의 기능을 한다. 즉, 환자가 진료를 받고 상급병원 진료를 위해서 자료를 요구할 때 진료기록 전송시스템에 요청하면 거점 문서저장소에서 자료를 협력병원이나 상급병원에 자료를 제공하게 된다. 개인정보에 대한 관리가 철저하게 이루어지는 가운데 진행이 된다.

### 3. 진료정보교류 현안 및 추진 방향

#### 가) 진료정보교류 현안

공공보건의료의 국가 책임과 지역격차 해소를 위해 권역·지역 책임의료기관 역할



을 강화하여 책임의료기관은 필수 의료(응급, 외상, 여성, 어린이, 재활, 장애인, 감염병 등) 서비스를 제공하고 권역 및 지역 책임의료기관 간 정보 전달 시 진료정보교류 시스템 활용을 권고하고 있다.

유관기관으로 언택트 기술 분야로 진료정보교류시스템 적용을 검토 중으로 중앙응급의료센터에서 중증 응급 환자 전원 시 응급실 진료기록의 전자적 제공을 위해 전남, 울산, 광주에서 10월 16일부터 시범 사업이 계획 중이다. 또한 근로복지공단의 산업재해 환자 관정에 필요한 의무기록·영상정보의 제출 방식·활용 방안을 검토 중이다. 병무청에서는 징병검사에 필요한 의무기록·진단서 등 전자적 교류 및 활용에 대해서도 검토 중이다.

한국판 뉴딜로 비대면 의료서비스 제공 등 스마트 의료 인프라를 확충 중으로 5G, IoT를 활용한 의료기관간 협진, 격리·집중치료실 등 비대면 의료서비스를 추진 중이며, 감염병 위험으로부터 환자·의료진 보호를 위해 진료정보교류가 대안으로 대두되었다.

국회에서는 의사 간 환자 진료기록교류 활성화를 위한 의료법 개정을 추진 중으로 의료인이 진료이력정보를 미리 확인할 수 있도록 개인 맞춤형 진료진원을 마련하고 있다.

## 나) 추진 방향

진료정보교류 범위 확대를 위해 의과 중심의 정보교류에서 치과, 한의과로 정보교류 범위를 확대를 위해 20년 서울대치과병원을 대상으로 시범사업을 추진하고 있다.

건강보험에서는 의료급여, 산재, 자바로 교류를 확대할 계획이다.

진료정보교류의 공공성 강화를 위해 중앙응급의료센터에서는 응급환자전원시스템 및 전원 협진망과 진료정보연계를 구축하고, 근로복지공단에서는 산업재해 관정 및 산재병원 간 환자 전원 시 정보교류체계를 구축하며, 보훈공단에서는 보훈대상자 관리 및 원 간 전원 시 정보교류체계를 구축하고, 심평원에서는 진료의뢰, 회송 시범사업 참여기관 간 정보연계를 확대할 계획이다.

의료와 복지정보 융합서비스 제공을 통해 임신·출산·사망 등 의료기관 정보 연계를 통한 임신부 및 영유아 등 복지서비스 수요 예측 및 사각지대 발굴이 가능할 것이다.

# 제 3차 포럼: 주제 2) 외국사례를 통해 본 진료정보 활용 방향과 실행 전략

발표자: 서울의대 박하영 교수

## 1. 의료체계와 진료정보 활용

국가 의료정보전략은 국가 의료체계 목표 달성을 지원해야 의미가 있다. 의료체계의 목표는 Access to care(접근성), Quality of care(health outcomes, effectiveness)(질), Cost of care(Efficiency)(비용, 효율성)이다. 의료 가치사슬은 생산자(산업), 제품 중간자, 의료 제공자, 지불자, 최종 사용자로 나뉘는데, 진료정보를 활용해서 할 수 있는 일들이 많이 있다. 생산자는 기술혁신을 위한 연구에, 의료 제공자나 지불자는 의료체계 접근성·질·효율성 향상을 위해, 구매자 및 사용자는 개인건강 예측·예방·관리 하는데 활용할 수 있다.

의료기관에서 데이터가 발생하면 이를 표준화하고 보호·보안된 정보를 저장하여 24시간 일주일 환자진료를 위한 정보 활용 및 정보 접근권을 가진 주체의 정보를 활용할 수 있는 시스템을 만드는 것이 중요하다. 또한 환자 및 건강한 개인이 건강관리를 위해 활용하고 교류할 수 있는 중앙 system을 갖추고, 국민건강을 모니터링하고 질병을 감시 하는데 활용되며, 약·의료기기 개발 등 연구개발 등에 활용되어 개발된 것이 의료에 적용이 되도록 하는 것이 중요하다. 정리하면 정보를 활용하는 것은 모여 있는 자료를 빅데이터로 활용하는 것, 그리고 진료나 건강관리를 하는데 활용되는 것으로 구성이 된다고 할 수 있다.

의료데이터 생태계는 의료데이터 제공자(의료인 및 의료기관, 보험자, 기타, 개인)와 이 데이터를 표준·호환성/정보보호·보완을 위한 의료데이터 인프라, 필요한 곳으로 흘러가도록 하는 게이트웨이, 잘 활용되도록 하는 산업 인프라, 최종적인 의료데이터 이용자 및 성과 등이 구성하고 있다. 즉 infrastructure layer, AI/Intelligence layer, Engagement layer 등 3개의 layer로 구성되어 있다.

## 2. 외국 사례

국가의 제공체계와 재정체계가 어떻게 되어있는지를 이해하는 것이 중요하다. 미국의 공보험(CMS)은 우리나라와 비슷하고 재정체계는 민간이 운영하며 이 두 가지가 같이

움직이며 운영된다. 영국은 제공체계, 재정체계 모두 공공으로 국가에서 예산을 배정해서 운영된다. 미국은 의료체계가 좋지 않기 때문에 이것을 해결하기 위해 정보 활용을 위해서 많은 노력을 하고 있다. 미국의 산업은 막강하고(제약, IT), 의료체계를 잘 활용하기 위한 시도를 많이 하고 있다. 의료제공체계 및 재정체계 혁신과 이를 지원하기 위한 의정보체계 인프라를 구축하고 있다.

미국 EMS EHR 인센티브 프로그램으로 민간의료기관들이 의료정보의 인프라를 만드는 투자를 인센티브 프로그램을 통해 진행하였다. 의료의 질, 안전성, 효율성을 개선하고 의료 격차(형평성)를 축소(개선)하며 환자와 환자가족의 진료참여를 촉진하고 의료기관 내, 의료기관 간의 진료 코디네이션을 개선하며, 국민의 건강과 공중보건을 개선하고 개인의 프라이버시와 정보의 안전성을 유지하기 위해 2009년부터 인센티브 프로그램을 시작하였다.

실행조직으로는 정책을 만들고 표준을 만드는 ONC와 실질적으로 집행하는 CMS가 있다. 성과관리 및 인센티브 지급 기준 지표는 1단계에서는 인증된 EHR 도입과 이의 의미 있는 사용, 2단계에서는 의료의 질 개선을 위해 의료정보를 교환하기 위한 인증된 EHR 기술을 이용, 3단계는 임상 질 지표를 포함한 정부가 선정한 관리지표들을 제출하는데 인증된 EHR 기술(전자적으로 호환 가능한 기술)을 이용하게 된다. 의료기관과 의료인의 자발적인 참여를 위해 2017년까지는 지원을 해주었고, 2017년 이후에는 패널티를 주는 구조로 프로그램이 짜여 있다.

지속적이고 일관성 있는 수정보환을 위해 2018년 4월 병원 프로그램 명칭을 Medicare Promoting Interoperability(PI) Program으로 변경하였고 2019년 Interoperability Rule을 발표하였다. 의료기관 시스템에 저장된 정보는 법적으로 환자의 소유임을 명확히 함으로써 환자의 건강기록 접근성을 보장하였고(PHR), 의료기관은 환자가 의료기관에 저장된 정보를 활용할 수 있도록 API를 개발해서 개방하도록 하였다. 현재 인센티브 프로그램은 본 사업으로 흡수된 상태이다.

의료기관이 확충되고 PHR의 개념이 도입되었다. 즉, 의료체계 문제 해결은 결국 개인이 지속적인 관리가 있어야 가능한데 이것이 잘 되지 않아 여러 가지 실패 사례가 있었다. 따라서 최근에는 호환 가능한 system으로 개인이 자기 건강을 관리하고, 건강기록을 쓰고 정부에서 이를 활용하는 형태로 진화하였다. 즉 의료기관의 EHR에 PHR을 얹는 형태이다.

영국은 여러 번 실패를 했고 NHS Digital이 2016년에 출범을 하여 진료정보 활용을 기대하고 있다.

### 3. 국내 시도

#### 1) 의뢰-회송 기반 진료정보교류 사업

2007-2010년 EHR 핵심 공동 기술 연구 개발 사업으로 분당서울대병원과 환자 의뢰-회송 협약이 있는 병의원들 간 의뢰-회송 환자 진료정보교류 시범사업을 하였다. 현재 보건복지부 진료정보교류 사업으로 2016년부터 상급종합병원의 신청에 따라 이들을 거점으로 연계된 1,2차 병·의원 간 진료정보교류를 시행하고 있으며 2017년 진료정보교류표준을 고시하였다.

#### 2) 4차산업혁명위원회 디지털헬스특위 개인 주도형 의료데이터 이용 활성화 전략

의료 데이터 이용을 활성화하자는 취지로, 예방·수요자 중심의 개인 의료데이터 연계·활용을 통한 국민건강증진 및 의료서비스 혁신, 그리고 이를 통한 의료체계 목표(환자중심 의료, 의료의 질 향상, 의료비 증가 억제) 달성을 목표로 하고 있다.

#### 3) 의료데이터 이용 활성화 개념도

산재되고 분절된 정보 연계를 통해 개인 주도적 의료데이터 통합·활용 하자는 것으로, 분절된 정보를 통합·연계하고 개인주도 의료데이터를 활용하여 개인 중심으로 모아서 활용하는 단기 및 중장기적으로 활용하자는 개념이다.

#### 4) 개인 주도형 의료데이터 실행 전략

어디 한군데에 정보를 모으지 않고 개인 의료데이터 보호·보안 및 개인 주도적 접근성·활용이 보장된 My Healthway라는 gate way를 구축하여 의료기관, 보험, 고용정보원, 산재 등 여러 가지 종류의 정보가 개인사용 동의를 하면 활용이 될 수 있도록 중심적인 역할을 할 수 있고, 정보가 필요할 때 언제든지 활용할 수 있도록 한다.

#### 5) 개인 주도형 의료데이터 중점 과제

개인 중심 의료데이터 통합·연계 체계 구축, 개인 정보 동의 및 보호 체계 구축, 개인중심 의료데이터 기반 서비스 발굴, 의료기관 및 국민 참여 활성화, 지속 가능한 혁신을 위한 거버넌스 등이 있다.

### 4. 진료정보 활용을 위한 전략

법제도는 개인정보 보호법도 통과가 되어 상당히 정리가 된 상태이다.

의료데이터의 표준화 부재로 인한 호환성 결여, 의료기관 의료데이터의 정보보호 및 보안 취약, 의료기관의 24/7 의료데이터 활용 여건 부재, 의료데이터가 발생하는 의료기관에서 의무기록으로 저장되는 정보의 내용 부실 등 의료기관 의료데이터 인프라 문제가 있다. 이를 위한 대안으로 인센티브 재원을 무엇으로 할지, 실행 주체를 누가 할지,

지금 방법은 어떻게 해야 할지에 대해 결정이 되어야 하는 상황이다. 의료데이터 이용 활성화를 위해 인프라가 구축되어야 할 것이고 이를 위해서는 투자가 필요하다.

### <질의 및 응답>

Q1. (전성연) : 우리나라에서 질지표 모니터링 자료를 의료기관에서 받는다면 진료정보 교류사업을 하고 있는 병의원들의 정보가 현재 표준화가 되어 있는지.

A1. (조경희) : EMR 업체에 표준화 하도록 지원하고 있다.

A2. (박하영) : 의뢰·회송하는 환자에 대해서는 표준화 된 것으로 알고 있다.

A3. (엄재선) : 의료 data는 부실한 상황이다. 전산에 대한 표준은 EMR로 cover가 가능하나, 의료 data 표준화는 용어 표준부터 시작해서 정리가 아직 안되어 있다. 사회보장정보원에서 용어 표준화에 대해 상급병원에서 mapping 작업을 진행 중으로 의료 data 표준화 작업을 하고 있는 중이다.

A4. (배철영) : 눈에 보이거나 눈에 보이지 않는 실제로 와 닿는 인센티브가 있어야 할 것이다.

A5. (유호범) : EMR 업체에 진료정보교류서식을 적용하는 모듈 개발을 지원하였고 현재 3종의 표준교류서식이 있고 이를 이용해서 진료정보를 교류한다. 영상은 별도로 첨부할 수 있다. 기본 정보를 입력하면 자동으로 기입되게 되어있다. 진료를 받고 기록할 때마다 진료요약지가 매번 생성이 된다. 자동으로 생성될지 수동으로 생성될지에 대해서는 아직 이슈가 되고 있다.

## 제 4차 포럼: 주제 1) 국가건강정보포털의 현황과 방향

발표자: 서울의대 조비룡 교수

### 1. 국가건강정보포털의 현황

국가건강정보포털은 국민의 수요에 기반하여 검증된 양질의 건강, 질병 정보를 통합적·체계적으로 제공하는 공공포털이라는 것에 의의가 있다. 핵심 가치는 검증된 정보(의학전문가의 광범위한 검증을 통해 신뢰할 수 있는 양질의 건강정보를 제공), 알기 쉬운 정보(쉬운 용어 사용, 시각자료 활용 등을 통해 일반 국민이 이해하기 쉬운 건강정보

제공), 통합적 정보(질병정보, 의료기관 정보, 의약품 정보 등 건강관련 정보의 통합·연계)이다.

국가건강정보 포털의 서비스 콘텐츠 현황은 현재 총 618건으로 질병 관련된 콘텐츠가 358개로 가장 많다. 유형분류로 보면 정신건강 관련 질병, 만성질환 자가 관리, 건강증진 생활습관 등에 관심이 많으나 현재 콘텐츠 수는 부족한 상황이다.

콘텐츠는 12개의 콘텐츠 유형을 선정하여 저자에게 제공하고 있으며, 콘텐츠의 일관성과 효율적 검색을 위해 유형별 집필 목차를 마련하였다. 감수는 문해력(literacy)을 보는 전문가들이 감수를 하게 되는데, 검증된 최신 지견, 국민들이 이해할 수 있는 내용으로 감수될 수 있도록 감수에 따른 매뉴얼을 제공하여 표준화된 감수를 할 수 있도록 하고 있다.

ontology란 각 단어가 어떤 맥락 속에 있는지 listing을 해서 질병의 진단, 치료, 관리가 엮일 수 있도록 만드는 것이고, 메타데이터는 정보를 누가 만들었는지, 언제 생성되었는지 관련된 내용을 보는 것이다. 이 두 가지가 한 집필 내용마다 붙여서 사용하게 된다. 저자들이 만들 때 ontology를 따로 만들지 않기 때문에, 반자동 과정을 수행하기 위하여 자연어 처리 기반 ontology 개발 및 관리를 하고 있다. 메타정보가 집필된 내용마다 들어가서 사이트가 업데이트 될 때 다른 정보들과 인터체인지가 가능하도록 만드는 것이 메타데이터이다. 이를 CMS(Content Management System)라고 하며, 네오젠에서 지원하고 있다. 온라인 교육 중 잘 알려진 것이 Learning management system(LMS)이며, LMS에서 콘텐츠를 제공하는 것이 Learning Contents Management System(LCMS)이고, 이 내용을 차용한 것이다. Learning management system은 필요한 것이 아니기 때문에 Content management system이 중요한 내용이라고 할 수 있다. 집필, 감수, 확인이 연계될 수 있도록 논문 심사 site를 많이 참조하였으며, 현재 업데이트 하고 있는 단계이다.

새로 개발한 사이트인 만성질환융합시스템은 작년까지 step 1의 단계였으며, 6개 시스템 데이터를 통합하였다. 올해는 4가지 시스템 데이터 통합 및 표준화 재개발을 진행하고 있으며, 여기에는 국가건강정보포털이 포함되어 있다. 시각화 할 수 있는 내용들, 역학, 통계자료들의 data 들이 국가정보포털로 바로 들어올 수 있도록 연계가 되는 과정이라고 볼 수 있다. 내년에는 질병관리본부에서 연구발주를 할 때 질병 관련된 데이터를 바로 쓸 수 있도록 하는 것이 만성질환융합시스템의 주요 내용이고, 국가건강정보포털의 정보들이 커넥션 될 수 있도록 하는 것이 주요 변화된 내용이다. 10월 7일부터 리뉴얼될 예정이다.

국가건강정보포털 운영체제는 질병관리본부, 운영위원회, 연구진, 내오젠 등으로 구성되어있다. 운영의 세부 내용으로는 건강정보포털 콘텐츠 개발(건강정보의 표준화, 콘텐츠 개발 지침 업데이트), 건강정보포털 운영 및 활용도 평가(상업적인 것이 아니라면 이용하도록), 건강정보포털운영위원회 사무국 운영, 운영위원회 사무국 역할, 집필진 워크샵, 집필진 인증제(어떻게 하면 쉽게 쓸지 연구) 등이 있다.

건강정보포털 운영 및 활용도 평가를 위해 2015년부터 구축된 '소비자 건강정보 모니터링 패널'을 확장 및 활용하고 있으며, 건강 문해력 및 활용도 평가 관련 조사, 소비자 패널단 추가 모집을 검토하고 있다. 또한 웹로그 분석(구글 또는 다른 분석도구)을 통해 가장 많이 검색한 콘텐츠와 유입경로, 유입 사이트에서의 입력 검색어, 정보 이용행태를 파악하여 소비자 건강정보 이용자를 분석하고 문해력에 대하여 정기적으로 평가하고 있다. 건강정보포털에서는 건강정보를 어떻게 만들 것인지 뿐만이 아니라 국민들의 literacy capacity 즉 어려운 정보를 어떻게 이해하도록 할지에 대한 것도 중요한 아이템 중의 하나이다. 건강정보포털 자문위원회에서는 건강정보의 제공자와 수요자가 진행을 하고 있다.

## 2. 국가건강정보포털의 발전방향

건강정보포털의 양대 축은 미국과 영국이다. 국민건강보험공단은 영국을 벤치마킹하면 좋을 것 같고, 질병관리본부는 Medlineplus를 벤치마킹하고 있다.

Medlineplus는 기존에는 정보를 모아두는 기능에서 search를 어떻게 해서 맞춤형 정보를 제공할 것인지로 바뀌고 있다. 내용이 매우 광범위하고 필요한 내용이 거의 다 들어있다. 따라서 search를 잘 하는 쪽으로 발전하고 있다. Medlineplus connect는 각 정보마다 질병코드, 처치코드가 tagging 되어있어서 나중에 국가 의료기관에서 사용하는 EMR, PHR에서 내 질병이 무엇이고, 자가 관리가 무엇인지 결정이 되면 코드가 tagging 되어서 그 코드와 맞는 정보들이 연계되어서 건강정보들이 EMR과 PHR에서 연동되도록 하는 것이 질병관리본부의 목표이다. 확장성과 맞춤형 정보가 들어가야 되므로 분절화하는 정보들을 ontology에 맞춰서 tagging하는 작업을 하고 있다.

영국의 NHS는 Literacy와 self care를 중요시 한다. 우리나라의 국민건강보험공단과 같은 것으로, 적절하게 의료를 사용하고 self care를 잘 해야 하므로, live well을 해야 하고 스스로 관리하고 자가진단을 할지와 관련이 되어있다. 스스로 자가 진단하는 tool이 있고, 이 과정에서 병원 예약까지 가능한 시스템이다.

호주와 캐나다는 이 둘의 중간적인 역할로 우리나라가 벤치마킹해도 좋을 것으로

생각된다.

해외국가건강정보포털의 주요 트렌드는 1. 방대하고 포괄적인 정보 탑재, 2. 검색 및 맞춤형 정보제공 기능 강화(ontology, DB structure), 3. 국민, 환자중심(그림, 동영상, literacy), 4. 최신정보제공(건강정보는 매년 20%가 바뀐), 5. 효과 증대 연계(outcome, 국민건강증진으로 이어져야 한다)이다.

국가정보포털의 문제점은 부분적인 건강정보 탑재, 검색 및 맞춤형 정보제공이 어려움(낙후된 DB 구조), 전문가 중심의 건강정보, update가 빠르지 않음, 가능한 자원 또는 실천프로그램과 연계가 많지 않은 점 등이 있다. 국가정보포털의 발전을 위해 국가차원의 체계적, 종합적 건강정보 제공 및 지속적인 질 관리가 필요하고 소비자의 요구에 부응하는 건강정보 제공체계의 구축 및 운영이 이루어져야 한다. 이를 위해 검색 및 맞춤형 건강정보를 제공(국가정보포털의 정보가 update가 되면 자료를 가져간 곳의 정보도 update되도록 하는 open API 도입), 국민 눈높이에 맞춘 건강정보 제공, 빠른 update를 통한 최신정보 제공, 효과증대를 위한 자원, 실천프로그램 연계 프로그램이 필요하다.

국가 건강정보포털 장기 운영 전략 개발을 위해 크게 세 개의 축이 필요하다. 건강정보 제공을 위한 거버넌스(인프라)를 구축해야하고, 국민들이 알기 쉽게 건강 정보를 제공하기 위해 이해하고 이를 진행해야하고, 국민들은 스스로 가짜정보와 진짜정보, 잘못된 정보를 구분할 수 있는 capacity와 역량을 강화할 수 있도록 해야 한다.

## <질의 및 응답>

Q1. A1.(조비룡) : 통계정보 및 매년 나오는 국건영 자료는 매년 이용해서 업데이트 될 수 있도록, 대국민 정보가 필요하면 정보를 가져갈 수 있도록 하고 있다. 내년부터 시행하도록 인프라를 구축할 계획이다.

Q2.(정중환) : 표준화가 된 것인지?

A2.(조비룡) : 각 질병마다 제공되는 정보가 다양한데, 필요한 정보를 가져갈 수 있도록 집필 시 tagging을 해서 검색할 수 있도록 정의, 역학, 증상 등을 세분해서, 전문용어를 쓰지 않고 한글로 집필하고 있다.

건강인에서 이것을 새로 개발하는 것 보다는 콘텐츠를 가져가는 것이 좋을 것으로 생각된다. 정보를 integration 할 수 있도록 질병관리본부와 협약을 맺는 것도 필요할



것으로 생각된다. 질병관리본부는 치료관련 부분은 약하고, 개발한 정보를 이용해서 질병관리에 대한 환자교육은 수가가 있어서 심사평가원에서 하고 있다.

건강인에서 업데이트 할 계획이 있다면 국립암센터의 암자료, 희귀질환, 심평원의 질병관리, 교육자료 등을 같이 쓸지에 대한 논의가 필요할 것으로 생각된다.

A2.(조정희) : 건강인의 약국정보 등과 국가 인프라를 같이 어떻게 합칠 것인지에 대한 논의가 필요하다.

## 제 4차 포럼: 주제 2) 비대면 환자교육 자료 소개

### 발표자: 헬스웨이브 정희두 대표

국가정보포털 콘텐츠는 대한의학회에서 13년 전에 만들었다. 소비자들의 눈높이가 이전보다 많이 높아졌고, 업그레이드 퀄리티도 높아졌다.

시각화된 애니메이션 콘텐츠를 만드는 것은 열 번 설명하는 것보다 한 번 보여주는 것이 효과가 뛰어나기 때문이다. EMR 자료에 인터넷에 있는 자료를 무단으로 잘못 쓰게 되면 문제가 되는 경우가 발생하는데, 직접 그러서 소유하는 방식은 비용도 많이 들고, 건강정보는 양이 너무 많기 때문에 하는 것도 불가능하고, 모든 의료기관마다 각자 하는 것도 예산낭비가 된다. SAS(Software as a Service) 같은 구독형 서비스처럼 CaaS(Contents as a Service)라는 월정액 서비스를 제공할 계획이다. 정부기관에서는 IP를 가지려고 하는데, 이는 정부가 무제한 무료 배포권을 가지고 있기 때문에 시장이 없어지게 된다.

현재 대형병원용 프리미엄 서비스와 중소병의원용 보급형 서비스를 제공하고 있다. 병원들이 공통으로 사용하는 멤버십 서비스로 운영되고 있으나, 표준화가 되어있지 않아 학회에서 검수하여 표준화된 콘텐츠를 공동개발 하는 것을 진행하고 있다.

### \*비대면 환자 교육

COVID-19 이후 서로 말을 할 수 없는 상황에서 untact 개념으로 원격으로 진료하는 상황에서 15분 진료에 어려움이 있을 수 있다. 따라서 설명 후 콘텐츠를 볼 수 있도록 콘텐츠를 활용하는 것이 좋을 것으로 생각된다.

스마트폰과 이메일로 맞춤형 애니메이션(설명처방)을 발송한다. 예를 들어서 감기환자에 대한 항생제 사용 설명에 대한 애니메이션을 환자에게 보내주면 이해하는데 도움이

된다. 개별병원이 만드는 것보다는 학회차원에서 만들면 객관적으로 받아들일 수 있는 자료가 될 것이다.

전공의 수 부족 등으로 동의서 업무가 많아지고 코로나 때문에 보호자를 병원으로 오게 하기도 어려운 상황이다. 설명의 효율성 등을 위해 학회차원에서 표준화된 작업이 필요하다. 애니메이션 내용을 글과 그림을 이용해 콘텐츠를 만드는 작업을 하고 있다. 현재 사용하는 전자동의서의 큰 문제점은 기존의 콘텐츠의 literacy에 대한 upgrade는 되지 않고 그대로 디지털 폼으로만 만든 것이다. 그림을 그리기도 어렵고, 인터넷 그림을 쓰다보면 저작권에 문제가 될 수도 있다. 동의서 내용을 알기 쉬운 애니메이션으로 만들고 텍스트와 그림을 서비스 회원 병원에 제공하고 내용을 채울 수 있도록 만들고 있다. 동의서 내용에 관한 애니메이션을 스마트폰으로 미리 보고 이후에 그 내용으로 동의서를 받는 방법 개발하고 있으며, 이를 이용하면 동의서 받는 시간도 줄일 수 있다. 비급여로 교육하고 수가를 받는 경우가 있으니 학회나 협회가 만든 표준 콘텐츠로 교육할 수 있도록 해야 한다.

수술 전후관리 교육상담 프로토콜에 퀴즈를 포함하여 환자가 교육을 읽었는지 분석할 수 있는 기획안을 만든 상태이다. 또한 환자에게 수술 관련 자료를 발송하고 수신했는지 확인할 수 있고, 몇분 동안 교육을 받았는지 교육이력 확인도 가능하도록 개발하였다.

신장학회에서는 복막투석에 대한 교육상담에 대한 수가를 만들었다. 환자에게 애니메이션을 미리 볼 수 있도록 하고, 대면교육용 웹슬라이드를 만들어 교육을 할 수 있다. 콘텐츠를 개발해서 병원에서 월 사용료를 지불하면 학회에도 지급이 된다.

당뇨학회, 고혈압학회, 한국지질동맥경화학회에서는 애니메이션을 개발하였다. healthbreeze 홈페이지에 데모시청 페이지가 있어서 토픽별로 데모 애니메이션 시청이 가능하다.

## 제 5차 포럼: 주제 1) 건강iN 방향 최종 토론회

주제: 개인건강기록 서비스 확대 방안-국민건강알림서비스 고도화 방안 등

발표자: 건강보험공단 빅데이터실 정종환 팀장

### 1. 앱을 통한 개인 맞춤형 서비스 제공 → 건강인 서비스 강화

- 1) (평생 건강정보) 공단의 건강관리 서비스 정보를 개인별 통합·연계하여 검진, 진단, 치료의 평생 건강정보 구축 → 청소년기 제외하고 영유아기, 20~30대, 40~50대, 노년기 등 연령별로 건강관리 서비스 제공
- 2) (맞춤형 건강서비스) 검진결과 등 건강정보에 기반 한 '개인의 생활 습관 및 질환 맞춤형 건강서비스'를 개발하고 적시에 제공 → 이를 확대하고 예측 정확도를 높이기 위해 연계 강화
- 3) (하이브리드 알람서비스) 그룹 단위의 특정 질환 알람 서비스와 감염성 질환 발생 시 개인 단위형태의 알람 정보 제공 → 국민건강알람서비스
  - 기후자료와 연계 미세먼지 자료와 연계, 대국민 지역별 위험 예측 정보를 보여줌
  - 감염병 및 계절성 질환 예측 계획
- 4) (질환발생 예측 고도화) 지식기반 질환발생 예측 모형을 고도화하고, 질환 발생률 감소를 위한 개인 맞춤형 건강관리 방법 제공
  - 위험 요인에 대한 개인 맞춤형 건강관리 서비스 제공

### 2. 신뢰성 있는 대표적인 건강·의료 플랫폼 제공

- 1) (건강정보) 전문가의 검증을 통한 신뢰성 있는 건강정보 확대
  - 건강인 매거진 매월 제작 중 → 일반적으로 보여주는 형태이므로 개인맞춤형으로 어떻게 보여줄 것인가 → 특정정보를 체크피 할 수 있는 건강수첩 기능 제공
- 2) (검진·진료예약) 병·의원 찾기 기능을 통하여 검진 및 진료 예약 가능 정보를 확인하고 예약할 수 있는 기반 구축
  - 검진률 높이기 위해 검진예약서비스 → 수검 예약 가능한 곳을 알려주는 기능 진행 고려 중
  - 진료예약은 장기적 과제
- 3) (진료정보교류) 병·의원 간 진료정보를 교류하여 의료이용자는 사전 동의로 상급

- 요양기관에서 1차기관의 진료정보를 볼 수 있는 기능 제공
- 4) (올바른 약물 이용) 약물 과다 복용자의 전문가 상담 기능 및 정보 제공
- 본인의 상담 정보를 PHR 차원에서 확인할 수 있는 연계 기능 제공
  - 건강인 포털사이트에서 복약 상담 → 이용률은 낮다.

### \*추가 내용

- 1) 데이터 개방 활성화에 맞춰서 공단에서 내년에 개인 건강기록 하나로 만들 계획
- 2) 데이터 댐
  - 보건의료 데이터를 한 곳에 모으려는 시도
  - 민간분야에 들어오지 않은 진료정보, 웨어러블 기기에 축적된 정보 → 국민에게 이로운 정보는 하나로 개인적인 서비스가 가능하도록 하는 방안을 마련하려고 함
  - 개인정보보호법에 의해 어떻게 보호, 어디까지 개방해야 하는가에 대한 법제화
  - 데이터가 활용 가능하도록 표준화 되어있으면 민간에서 활용할 수 있을 것

### \* 논의 내용

- 진료정보 교류 : 어떻게 이용할 수 있을지
- 검진 예약 시스템 : 검진청구시스템 이용하는 소규모 병원, 포털 시스템에서 예약 가능한 병원 안내 가능하도록
- 질환발생 예측 고도화 : 질환 발생 예측을 해서 무엇을 해 줄 것인가에 대한 목표를 정하면 모델을 만드는데 도움이 될 것으로 생각 된다 ↔ 암환자도 생활습관 교정해서 좋아지는 evidence가 있다, 단기적인 효과가 없더라도 생활습관 교정, 예방 필요, 예방에 공단이 나서서 것이 좋겠다.
- 약물 이용 업그레이드 필요함

# 제 5차 포럼: 주제 2) 건강iN 연구보고서(2019년)

발표자: 메디에이지 임유리 수석연구원

## 1. 연구목적

- 건강iN 사이트의 역할 정립
- 건강iN 사이트(웹·앱)에서 제공하는 서비스 전반에 대한 개선 방안 제시
- 모바일 특화 콘텐츠 개발 및 실행방안과 메뉴 재구성(안) 제시
- 맞춤형 건강저오 콘텐츠 [건강나이] 서비스 제공 방안 마련
- 맞춤형 개인별 건강서비스 제공을 위한 단계별 로드맵 및 비전 제시

## 2. 건강iN 사이트의 역할 정립

- 국민에게 확실한 신뢰를 얻을수 있는 개인건강데이터를 활용한 맞춤형 정보제공
- 건강·의학정보의 제공 뿐 아니라 개인 건강검진 결과에 기반한 맞춤형 건강정보 및 건강관리 프로그램 제공
- 신뢰성 있는 건강정보 제공을 기본으로 개인 맞춤형 건강정보 제공 및 건강관리 도구로서의 역할에 초점을 맞춘 고도화

## 3. 건강iN 사이트(웹·앱)에서 제공하는 건강정보 제공서비스 전반에 대한 콘텐츠 진단 및 분석

- 1) 국내 공공, 민간 기관 분석, 국내 공공기관 분석
- 2) 웹사이트 개편 방안
  - 서비스와 정보를 쉽게 찾도록 함
  - 서비스의 사용 편의성을 높임
  - 신뢰할 수 있는 정보와 서비스 제공
  - 서비스를 안정적으로 이용할 수 있도록 함
  - 보안과 개인정보 유출에 대한 우려 없이 안심하고 사용할 수 있도록 함
  - 건강상태 측정 편의성
  - 복약 알람 기능 개선

4. 모바일 앱 건강iN을 통해 제공가능한 모바일 특화 콘텐츠 개발
5. 맞춤형 건강정보 콘텐츠 [건강나이] 서비스 제공 방안 마련
6. 맞춤형 개인별 건강서비스 제공을 위한 단계별 로드맵 및 비전 제시  
 통합→특화→안정화→확산→ [건강iN 비전 2022] “국민 의료비 지출을 감소에 기여하는 빅데이터 기반 맞춤형 건강관리 사이트”

**\* 논의 내용**

- 1) 건강iN 사이트 로그인 시 개인이 본인의 정보를 다 볼 수 있도록 해줘야 하고, 데이터 댐 구축하여 개인의 건강정보 볼 수 있도록 구축하고 논의 되어야 함.
- 2) DB 구축 시 고유식별번호 필요함 → 만성질환별로, 건강검진 별로 따로 있음.

## 제 5차 포럼: 주제 3) 2021년 연구 관련

**발표자: 조경희 교수**

### 1. 국가검진 표준 연구

- 매해 검진→ 국민의 생활습관을 구분해서 만들어 놓음. BMI, 콜레스테롤, 대사증후군 등이 매년 어떻게 바뀐다는 것을 볼 수 있도록 시스템을 만들어 놓아야 함.
- 공단 PHR에서 연계되어 스스로의 건강 수준을 알 수 있도록 시스템 필요

### 2. 고령화 사회를 맞이하여 Age related disease (노화관련 질환) 발생률 및 질병 예측

- 치매, 전립선질환, 골다공증, 관절염, 백내장, 황반변성 등

연구보고서 2020-20-018

## 개인별 맞춤형 통합 건강관리 프로그램 연구 및 개발(7차년도)

---

발행일	2021년 2월 5일
발행인	김성우
편집인	오현철
발행처	국민건강보험 일산병원 연구소
주소	경기도 고양시 일산동구 일산로 100(국민건강보험 일산병원)
전화	031) 900-6977, 6985
팩스	0303-3448-7105~7
인쇄처	백석기획 (031-903-9979)

---



(우)10444 경기도 고양시 일산동구 일산로 100(백석1동 1232번지)  
대표전화 1577-0013 / 팩스 031-900-0049  
[www.nhimc.or.kr](http://www.nhimc.or.kr)

## 2020 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