

과거질병이력과 건강검진지표에 기반한 치매 발병 예측모형 개발 - 국민건강보험공단 노인코호트DB 사용

전희주*, 인태교**

국문초록

본 연구는 건강보험공단의 노인코호트DB를 사용하여 개인의 건강검진 데이터와 과거질병이력 데이터의 치매상병코드를 이용해 통계적 예측모형을 개발하여 우리나라 노인들의 객관적인 치매 발병 요인을 분석하였다. 노인 치매 발병에 가장 큰 요인인 연령을 제외하고 치매 발병에 영향을 주로 미치는 요인은 최근 5년 동안의 질병 발생 여부이다. 최근 5년 동안 뇌질환, 파킨슨병 유병 여부, 7일 이상 계속 치료, 입원, 당뇨병, 골절 유병, 허혈성심장질환 유병, 수술, 고혈압 등이 있을수록 노인 치매 발병률이 높아지는 경향을 보였다. 다음으로 건강검진 지표들이 노인 치매 발병에 영향을 주는 중요한 요인이었다. 감마지티피, 혈색소, 공복혈당, 총콜레스테롤, 수축기혈압, 이완기혈압 등 건강 상태가 치매 발병에 영향을 주는 중요한 지표임을 함께 보였다.

본 연구에서 제시된 치매 발병률 추정값은 치매 리스크에 노출될 수 있는 미래 치매 환자들과 그 가족들의 경제적 리스크를 준비할 수 있는 보험상품 개발에 활용되고 개인별 과거 병력 및 건강 상태에 따라 위험을 세분화할 수 있을 것이다. 또한 노인코호트DB를 사용하여 실제 노인들의 과거 질병 이력과 건강검진 지표들을 통해 치매발병 상대위험도 예측 모델의 개발을 통해 치매 예방 연구에 새로운 방향을 제시하고자 한다.

□ 주제어: 건강검진, 과거질병이력, 노인코호트DB, 로지스틱회귀모형, 치매발병

□ JEL: J23

투고일: 2024.10.16.

수정일: 1차 2024.12.2., 2차 2025.1.12.

게재확정일: 2025. 4.11.

* 동덕여자대학교 정보통계학과 부교수(hjchun@dongduk.ac.kr), 제1저자

** 한양대학교 보험수리학과 박사(tkleen@naver.com), 교신저자

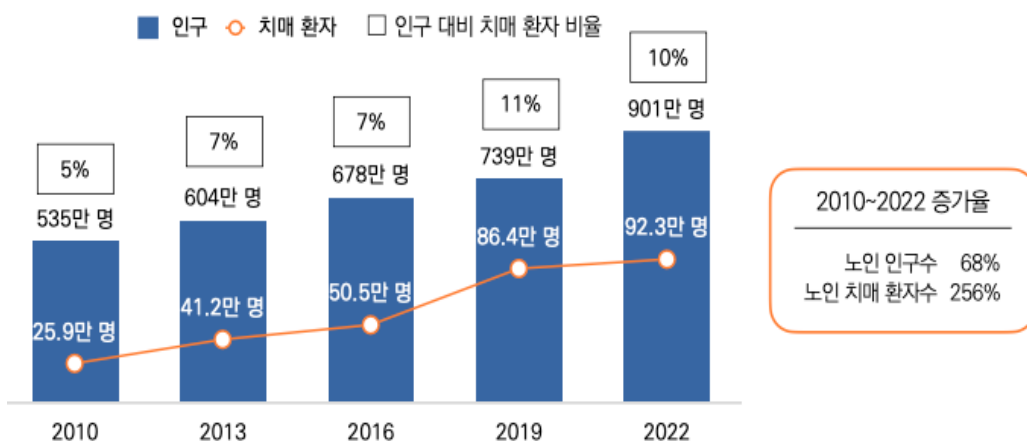
I. 서론

치매는 전 세계적으로 심각한 공중보건 문제로 부상하고 있으며, 특히 인구 고령화가 급속히 진행되는 국가들에서 그 중요성이 더욱 강조되고 있다. 치매는 주로 노년층에서 발병하지만, 발병 원인과 위험 요인은 매우 다양하다. 치매의 발병을 예측하고 예방하는 것은 개인의 삶의 질을 높이고, 사회적 및 경제적 부담을 줄이는 데 중요한 역할을 한다.

2023년 12월 말 대한민국 65세 이상 고령인구는 973만명으로 대한민국 전체 인구의 19.0%(통계청, 2024)로 65세 이상이 20%로 정의되는 초고령사회 진입을 눈앞에 두고 있다.

중앙치매센터에서 발표한 2023년 대한민국 치매 현황에 따르면, 2022년말 기준 65세 이상 노인 인구 중 치매 환자 수가 약 92.3만 명으로 65세 이상 노인의 10.2%가 치매를 앓고 있는 것으로 나타났다(국립중앙의료원, 2024).

[그림 1] 65세 이상 노인 치매 환자수 추이



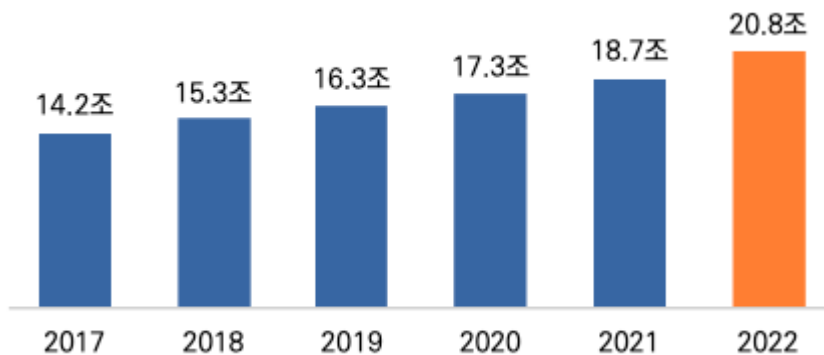
(출처: 국립중앙의료원 중앙치매센터, 2024)

향후 65세 이상 노인 치매 환자 수는 고령화가 급속히 진행됨에 따라 지속적으로 증가하여 2040년 226만명, 2050년에는 315만명까지 증가하여 2023년 현재의 3배 이상이 될 것으로 추정된다(국립중앙의료원, 2024). 65세 이상 노인 인구 중 치매 환자 비율은 2023년 10.4%

에서 2050년에는 17%까지 크게 오를 것으로 예측된다.

2022년 치매 환자 1인당 연간 관리 비용은 2,220만원, 65세 이상 치매 관리 비용은 약 22조 6,467억 원에 달해, 치매가 사회적 및 경제적 측면에서 큰 부담이 되고 있다. 2022년 국가 치매 관리 비용은 20.8조원으로 추정되며 향후 치매 환자 수는 지속적으로 증가하고 있어(국립중앙의료원, 2024), 국가 치매 관리 비용 역시 크게 증가할 것으로 예측된다. 이러한 상황에서 치매 발병 가능성을 예측하고 예방하는 연구는 더욱 중요해지고 있다.

[그림 2] 65세 이상 국가 치매 관리 비용(단위: 원)



(출처: 보건복지부, 중앙치매센터, 2024)

기존 치매 발병 예측 연구는 주로 인체 자원(혈액, DNA 등), 임상 정보, 뇌 영상 정보(MRI, CT 등)를 수집하고 분석하는 방식으로 이루어졌다. 이러한 접근법은 매우 정밀한 데이터를 제공하지만, 일상적으로 접근하기 어렵고 비용이 많이 든다는 단점이 있다. 이에 따라, 본 연구에서는 일상적으로 흔히 접할 수 있는 건강검진 정보와 과거 질병 이력을 활용하여 예측 모델을 개발하고자 한다. 이러한 접근법은 보다 쉽고 경제적으로 치매 위험을 평가할 수 있는 장점을 지닌다.

특히, 건강검진 데이터는 주기적으로 갱신되며, 다양하고 필수적인 건강 지표를 포함하고 있어 치매 예측 모델의 정확성을 높이는 데 중요한 역할을 한다. 혈압, 혈당, 콜레스테롤 수치 등의 생물학적 지표와 함께, 흡연, 음주 등 생활 습관에 관한 정보가 포함되어 있어, 종합적인 건강 상태를 평가할 수 있다. 과거 질병 이력 역시 치매 발병과 밀접한 연관이 있는 중요한 데이터로써, 고혈압, 당뇨병, 심혈관 질환 등의 기존 질병이 치매 위험을 증가시키는 요인으로 작용할 수 있다.

현재 한국의 보험사들이 사용하는 치매 관련 위험률은 국내 관련 데이터를 충분히 활용하지 못하고 국가에서 공표하는 치매 유병률을 이용하여 치매 발생 및 사망 비율을 일본의 방법론을 참고하여 산출하는 실정이다. 향후 고령화가 진전됨에 따라 치매 발병 정도는 건강상태에 따라 차이가 증가되는 상황에서 이에 대한 위험 요인을 판별하고 이를 개인별로 식별하고 분석함으로써 개인 맞춤형 예방 전략 수립이 필요한 시점이다. 특히, 치매는 그 특성상 고령으로 진행되면서 급격하게 발생하는 질병으로써 장기적 관점에서 치매 위험을 관리하지 않으면 장기적으로 치매 환자가 증가함에 따라 손해율이 상승할 우려가 존재하며, 보험사의 사회적 기능을 위해서도 보험사는 보험계약자의 건강관리를 통해 위험 요인을 관리하는 것이 중요해지게 될 것이다. 따라서 현 시점에서 우리나라 실정에 맞는 건강상태별 치매에 영향을 미치는 주요 요인의 위험도를 측정할 수 있는 예측 모델 개발이 필요하다는 것을 의미한다.

이제는 국가에서 관리중인 공공데이터가 점차 개방되면서 한국 실정에 맞는 데이터를 활용할 수 있는 기회가 확대되고 있어, 국내 데이터를 기반으로 과거 병력 및 건강상태 등 치매에 영향을 미치는 요인을 보다 정확히 반영한 위험도 분석이 가능한 상태에 있다. 향후, 고령화로 인해 치매 발병률이 급증할 수 있는 상황에서 국내 데이터를 활용한 예측 모델을 개발하는 것은 보험사의 리스크 관리와 재무 건전성 유지에 있어 중요한 역할을 할 수 있으며, 보험사의 장기적인 안정성에도 기여할 것이다.

본 연구의 목적은 건강검진 데이터와 과거 질병 이력 데이터를 바탕으로 치매 발병 가능성을 예측하는 모델을 개발하는 것으로써, 이를 위해 국민건강보험공단의 노인코호트DB를 활용하여 통계적 예측모형을 개발하고, 통계 기법을 적용하여 데이터의 패턴을 분석함으로써, 보다 합리적인 알고리즘을 도출하고자 한다. 이러한 예측 모델은 개인 맞춤형 예방 전략 수립과 조기 발견을 통해 치매로 인한 사회적, 경제적 부담 및 비용을 경감시키는 데 기여할 수 있을 것이다.

본 연구는 우리나라 객관적인 치매발병의 위험요인을 얻기 위해 국민건강보험공단이 노인 코호트DB를 가지고 개인의 건강검진 데이터를 바탕으로 통계적 예측모형을 개발하고 발병 위험도를 예측하고자 한다. 또한, 본 연구에서 제안된 치매 발병 예측모델은 개인의 건강 데이터를 종합적으로 분석하여 치매의 조기 발견과 예방 연구에 중요한 도구가 될 수 있으며, 예측된 치매발병률은 치매보험 개발을 위한 중요한 기초자료로 활용될 수 있을 것이다. 결론적으로 본 연구의 치매 발병률 예측모델 개발과 검증을 통해 새로운 방향을 제시하고, 공중보건 향상에 기여하고자 한다.

II. 이론적 배경 및 선행연구 분석

치매는 후천적인 외상이나 질병 등 외부요인에 의하여 정상적으로 성장한 뇌가 손상 또는 파괴되어 전반적으로 지능, 학습, 언어 등의 인지기능과 정신기능이 떨어지는 복합적인 증상의 노년기 대표 정신장애이다. 치매는 중요한 신경 질환으로 현재 심장병, 암, 뇌졸중과 함께 4대 주요 사망원인 중의 하나이다(질병관리청, 2024).

치매는 한번 발병하면 완치나 중단시킬 수 있는 그 어떤 의료적 처치나 처방은 없으나, 예방 및 조기발견으로 노력에 따라 진행속도를 늦추거나 호전을 기대할 수 있는 질환이다(Navrátlová et al., 2007). 치매는 연령이 증가하면서 치매의 발생은 급속히 증가하는데, 65세 이후 매 5년마다 두 배씩 증가한다고 알려져 있다(John, 1991).

치매의 원인 질환으로는 80~90가지가 알려져 있지만 그중에서 주요 3대 원인 질환은 ‘알츠하이머병’, ‘혈관성 치매’ 등이 있다. 알츠하이머병은 전체 원인의 약 50%를 차지하여 가장 흔한 치매의 원인이다. 뇌졸중 후에 동반되는 혈관성 치매는 약 10-15%이고, 알츠하이머병과 혈관성 치매가 동시에 발생하는 경우는 약 15%로 나타나고 있다(질병관리청, 2024). 기타 질병에 의한 치매로는 ‘픽병’, ‘크로이츠펔트-야콥 병’, ‘헌팅톤 병’, 그리고 ‘후천성 면역 결핍증에 의한 치매’, ‘알코올성 치매’ 및 ‘뇌 손상 후의 치매’가 있다.

알츠하이머성 치매는 시간이 경과함에 따라 질환이 지속적으로 진행되는 특징이 있기에 진행성(퇴행성, 비가역성) 치매로 분류하는 반면, 혈관성 치매는 뇌혈관 질환에 따라 치매발생이 달라지므로 예방이 가능한 치매로 분류된다(한수정 · 백남중, 2015).

현재 치매 치료에 있어 가장 대표적인 방법은 약물치료이다. 치매 환자 약물치료의 목표는 질환의 진행 속도를 최대한 늦추는데 초점이 맞추어지는데, 이는 아직까지 치매 치료제로써 질환 경과의 방향을 전환시키고 뇌 기능을 회복시키는 약제가 개발되지 못하였기 때문이다.

또한 아직까지 치매의 근본적인 원인은 밝혀지지 않고 오로지 예방하거나 약물치료로 치매의 진행을 늦추는 방법밖에 없는 실정이다. 그러나 노인 치매 증가의 58%가 긴 수면시간과 연관이 있고(Benito-León et al., 2014), 수면부족과 수면의 질이 치매와 관계가 있다는 연구도 다수 존재한다(Henry et al., 2019; Leng et al., 2019).

국내에는 현재까지 과거 질병이력과 건강검진에 근거한 치매발생률에 대한 연구는 없다. 치

매발생률을 예측한 선행연구로는 보험수리적 방법(권혁성, 2012), 치매진단과 수면관련 데이터를 통한 머신러닝 방법(조영은 외, 2024), 3D 이미지 국외데이터를 이용한 머신러닝 방법(이태인 · 오하영, 2021)이 있고 보건복지부에서 주관한 2016년 전국 치매역학 조사가 있다. 2016년 치매역학 조사는 전국 17 개 광역시도를 모두 포함한 대규모 표본 역학 조사로서, 국제적으로 사례가 드문 정부 주도의 유병률 조사 연구이다.

권혁성(2012)은 실제 데이터가 아닌 보건복지부 자료를 이용하여 민영 보험의 수요 추정을 위해 미래 특정 년도의 인구 추계를 바탕으로 치매 환자 수를 추정하는 모형의 설계 방법을 보험수리적으로 제안하였고 성별, 연령별 치매 유병률과 발병률의 추정치를 제시하였다.

조영은 외(2024)는 174명을 대상으로 웨어러블 기기에서 수집된 라이프 로그 정보와 치매진단 데이터를 활용하여 머신러닝에 기반한 치매예측모형을 개발하였다. 이태인 · 오하영(2021)은 Marcus et al.(2007)의 OASIS-3 이미지 데이터를 머신러닝 기법을 이용하여 분석하고 치매를 예측하는 모형을 제안하였다. 그러나 조영은 외(2024)와 이태인 · 오하영(2021)이 제안한 치매발병 예측모형은 치매예측을 위한 머신러닝을 적용할 수 있는 방법론을 제시했지만 보험사나 보험 정책적인 면에서 실질적으로 사용할 수 있는 치매발병률 예측값의 형태는 아니다.

김민국(2019)은 국내외 치매보험시장과 일본의 치매정책과 치매보험시장 현황을 국내의 시장 상황과 비교 및 분석하였다. 조맹제(2009)는 2008년 치매 노인 유병률 조사의 결과를 바탕으로 국내 치매노인 유병률 현황과 위험 요인의 연구주제로 알츠하이머형 치매와 혈관성치매의 위험 인자를 분석하였고, 치매 환자의 조호 비용과 부담을 분석하여 이와 깊이 관련된 치매 유병률을 낮출 수 있는 방안을 제시하였다.

III. 연구설계 및 연구 방법

1. 분석데이터와 연구설계

본 연구에서 활용하는 데이터인 국민건강보험공단의 노인코호트DB는 2008년 1월 기준 건강보험 및 의료급여 자격을 유지하고 있는 만 60세 이상의 약 640만명 중 8%인 약 51.2만명을 성, 연령(1세단위), 지역구분(대/중/농어촌), 보험료(10분위)에 의한 층화무작위추출을 하여 2002~2019년의 자격 및 사회·경제적 정보, 병·의원 이용내역 및 건강검진 결과, 요양기관 정보와 노인장기요양서비스 신청 및 이용 내역(2008~2019) 등을 전향적 + 후향적 코호트 DB입니다.

본 연구는 노인코호트DB에 대한 IRB(Institutional Review Board, 의학연구윤리심의위원회) 승인을 받아 국민건강보험 가상화DB에서 모든 분석을 진행하였다.

본 연구에서 치매발병예측을 위한 연구설계는 [그림 3]과 같이 2015년 1월을 기준년월로 설정하였고 치매원인을 찾는 관찰기간(observation period)은 기준년월 기준 과거 5년 2010~2014년으로, 치매발병자를 측정하는 평가기간(performance period)은 2015년~2019년으로 설정하였다. 분석대상자는 2015년 1월 기준년월 기준 최근 2년 2013~2014년 동안에 건강검진 기록이 있고 과거 5년 2010~2014년 동안 치매진료이력이 없는 노인을 대상으로 하였다. 보험회사에서 인수심사(U/W)에 활용하는 고지항목(계약 전 알릴 사항)에 과거 5년간 의료이력을 묻고 있는 실정을 고려하였고, 건강검진 기록의 경우 가장 최근 검진 결과를 활용하고자 하였으며 일반건강검진 주기가 2년임을 고려하였습니다.

2015년 1월 기준 중도 탈퇴자 및 사망자, 과거 5년인 2010~2014년 진료내역 기록이 없는 의료급여 수급권자는 분석대상에서 제외하였다. 본 연구는 5년 뒤 특정 시점의 치매발생 여부를 파악하는 것이 아니라 5년 이내 특정 기간 동안의 치매발생 여부를 파악하는 것이므로 치매평가 기간(5년) 중 사망 발생 건 또한 포함하여 분석하였다. 치매이력 및 치매발병 유무는 진료내역 중 주상병과 부상병의 치매상병코드(F00, F01, F02, F03, F051, G30)인 경우로 설정하였다.

[그림 3] 치매예측모형 관찰기간과 평가기간 정의



[표 1]은 본 연구에서 관찰기간 5년 동안의 관찰기간의 질병이력과 치매 정의를 하기 위한 KCD코드와 진료행위코드이다.

[표 1] KCD코드와 진료행위코드

구 분		KCD 상병코드
상병명	치매	F00~F03, F051, G30
	암	C00~C97, D45, D46, D471, D473, D474, D475
	뇌혈관질환	I60~I69
	허혈성심장질환	I20~I25
	파킨슨병	G20
	골절	S02, S07, S099, S12, S22, S32, S42, S52, S62, S72, S82, S92, T02, T08, T10, T12, T142
	고혈압	I10~I13, I15, I674, O10, O11, O13~O16, H350
	고지혈증	E78
진료 행위	당뇨병	E10~E14, E748, E891, G590, G632, G990, H280, H360, M142, N083, O24, P702, R730, R81
	수술	OPRTN_YN(수술여부) = 9 0: 미수술, 9: 수술
	입원	FORM_CD(서식코드) = 02, 04, 06, 07, 10, 12 02: 의과입원, 03: 의과외래, 04: 치과입원, 05: 치과외래, 06: 조산원입원, 07: 보건기관입원, 08: 보건기관외래, 09: 정신과낮병동, 10: 정신과입원, 11: 정신과외래, 12: 한방 입원, 13: 한방기관외래, 20: 약국조제, 21: 처방조제

[표 2]는 분석대상자들의 인구통계적 특성이다. 분석대상자 중 남성의 비율은 46.52%, 여성의 비율은 53.48%이다. 연령대를 보면 55~60세 미만 26.02%, 60~65세 미만 25.59%, 65~70세 미만 19.98%, 70~75세 미만 15.55%, 75~80세 미만 9.2%, 80~85세 미만은 3.46%이고, 85세 이상은 1.2%이다.

장기요양급여형태는 장기요양급여를 받지 않은 급여미대상이 99.64%로 대다수이고, 장애 중증정도는 해당이 없는 경우가 89.55%로 대다수를 차지하고 있다. 소득분위는 0분위에서 10분위까지 있으며 분위가 높을수록 소득이 높음을 의미한다.

[표 2] 분석대상자의 인구통계적 특성

(단위: 명, %)

변 수		대상자수	%
성별	남성	272,487	46.52
	여성	313,258	53.48
연령	55세 ~ 60세 미만	152,387	26.02
	60세 ~ 65세 미만	149,890	25.59
	65세 ~ 70세 미만	117,013	19.98
	70세 ~ 75세 미만	85,229	15.55
	75세 ~ 80세 미만	53,887	9.20
	80세 ~ 85세 미만	20,291	3.46
	85세 이상	7,048	1.20
장기요양급여형태	급여미대상	583,652	99.64
	시설급여	323	0.06
	재가급여	1,717	0.29
	재가시설급여	53	0.01
장애중증	해당없음	524,527	89.55
	중증(1~2등급)	14,626	2.50
	경증(3~6등급)	46,574	7.95
소득분위	의료급여수급권자	18,567	3.17
	1분위	50,261	8.58
	2분위	40,648	6.94
	3분위	44,350	7.57
	4분위	37,003	6.32
	5분위	41,552	7.09
	6분위	49,177	8.40
	7분위	57,727	9.86
	8분위	68,104	11.63
	9분위	80,892	13.81
	10분위	97,446	16.64
계		585,745	100.00

[표 3]은 분석대상자들의 과거 5년 동안의 질병이력 특성이다. 최근 5년 동안 수술을 경험한 노인들은 69.85%, 최근 5년 동안 입원을 한 대상은 50.59%, 최근 5년 동안 7일 이상 계속 치료를 받는 경우는 61.15%로 나타났다. 최근 5년 동안 질병 이력이 있는 경우는, 암 5.99%, 뇌혈관질환 11.72%, 허혈성심장질환 13.84%, 파킨슨병 0.53%, 골절 18.08%, 고혈압 51.26%, 고지혈증 44.64%, 당뇨병 29.26%로 나타났다.

[표 3] 표본의 과거 질병이력 특성

과거 질병이력		대상자수	%
최근 5년 수술 여부	있음	375,832	69.85
	없음	208,869	30.15
최근 5년 입원 여부	있음	293,421	50.59
	없음	292,324	49.91
최근 5년 7일 이상 계속치료 여부	있음	358,158	61.15
	없음	227,587	38.85
최근 5년 암 유병 여부	있음	35,106	5.99
	없음	550,639	94.01
최근 5년 뇌혈관질환 유병 여부	있음	68,634	11.72
	없음	517,111	88.28
최근 5년 허혈성심장질환 유병 여부	있음	81,071	13.84
	없음	504,674	86.16
최근 5년 파킨슨병 유병 여부	있음	3,111	0.53
	없음	582,634	99.47
최근 5년 골절 유병 여부	있음	105,874	18.08
	없음	479,871	81.92
최근 5년 고혈압 유병 여부	있음	300,269	51.26
	없음	285,476	48.74
최근 5년 고지혈증 유병 여부	있음	261,465	44.64
	없음	324,280	55.36
최근 5년 당뇨병 유병 여부	있음	171,415	29.26
	없음	414,330	70.74
계		585,745	100.00

[표 4]는 분석대상자의 건강검진 특성 변수들에 대한 기초통계량이다. 최근 5년 동안 수술 횟수의 평균은 1.54회이고, 최근 5년 동안 입원일수의 평균은 12.5일이다.

[표 4] 분석대상자의 건강검진 특성

변수	영문(약자) 코드	대상자수	평균	표준편차	최솟값	최댓값
연령	age	585,745	65.50	7.31	56.00	87.00
최근 5년 수술횟수	SUR_TIMES_5Y	584,701	1.54	2.07	0.00	163.00
최근 5년 입원일수	INPATIENT_DAYS_5Y	585,745	12.50	54.90	0.00	1827.00
체질량지수(kg/m) ²	BMI	585,631	24.14	3.06	11.20	70.30
수축기혈압(mmHg)	BP_HIGH	585,675	126.12	14.94	60.00	265.00
이완기혈압(mmHg)	BP_LWST	585,675	77.02	9.65	35.00	180.00
공복혈당(mg/dL)	BLDS	585,990	104.00	26.64	12.00	951.00
총콜레스테롤(mg/dL)	TOT_CHOLE	585,990	196.56	41.30	12.00	2655.00
중성지방(mg/dL)	TRIGLYCERIDE	585,688	134.81	87.25	1.00	5560.00
HDL콜레스테롤(mg/dL)	HDL_CHOLE	585,687	53.71	16.07	1.00	3915.00
LDL콜레스테롤(mg/dL)	LDL_CHOLE	583,590	116.39	43.55	1.00	9999.00
혈색소(헤모글로빈, g/dL)	HMG	585,688	13.75	1.46	0.80	24.50
혈청크레아티닌(g/dL)	CREATININE	585,689	0.90	0.73	0.00	98.00
혈청GOT(ALT, U/L)	SGOT_AST	585,693	26.95	20.45	1.00	7120.00
혈청GPT(ALT, U/L)	SGPT_ALT	585,683	24.51	20.55	1.00	2957.00
감마지티피(U/L)	GAMMA_GTP	585,691	36.07	52.07	1.00	999.00

[표 5]는 2015년 1월 1일 기준(분석기준 시점) 과거 5년 동안 치매발병이 없고 최근 2년 동안 건강검진을 받은 55세 이상인 분석대상자 585,745명 중 분석기준 시점을 기준으로 할 때, 향후 5년 동안 치매가 발생한 대상은 34,993명 5.97%로 나타났다.

[표 5] 분석대상자의 치매발생률

(단위: 명, %)

		대상자 수	비중
치매발생	없음	550,752	94.03
	있음	34,993	5.97
계		585,745	100.00

2. 통계적 예측모형

본 연구는 55세 이상의 고령자들을 대상으로 치매 발병률을 예측하기 위해 다중 로지스틱 회귀모형(multiple logistic regression model)을 활용하였다. 다중 로지스틱회귀모형은 일반화선형모형(generalized linear model)중 반응변수가 이항분포를 따르는 경우에 해당되며, 반응변수가 이항인 경우 영향요인들을 통해 발생률을 예측하고자 할 때 사용되는 통계적 모형이다(Agresti, 2013). 다중 로지스틱회귀모형은 아래의 식 (1)과 같이 표현된다.

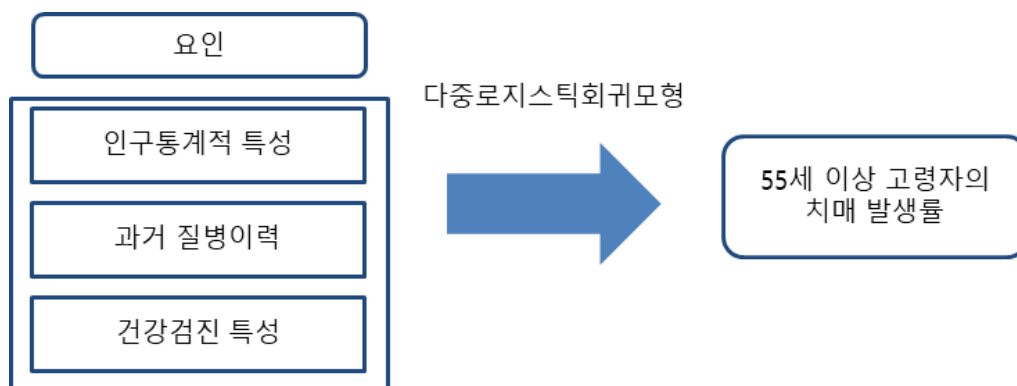
$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k \quad (1)$$

여기서 π 는 55세 이상 고령자들의 치매발생률을 나타내고 $x_1, x_2, \cdots, x_k$ 는 노인들의 치매에 영향을 주는 요인들이다. 또한, 치매발생률에 대한 추정값 $\hat{\pi}$ 은 식 (1)을 통해 얻어진 식 (2)로부터 구해진다.

$$\hat{\pi} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1x_1 + \hat{\beta}_2x_2 + \cdots + \hat{\beta}_kx_k)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1x_1 + \hat{\beta}_2x_2 + \cdots + \hat{\beta}_kx_k)} \quad (2)$$

본 연구의 55세 이상의 고령자들을 대상으로 치매 발병률을 예측 모형은 [그림 4]와 같이 도식화하여 표현할 수 있다.

[그림 4] 치매예측 연구모형



IV. 분석결과

분석대상자들의 인구통계적 특성, 최근 5년 과거 질병이력, 건강검진 특성 변수들을 사용하여 고려된 35개의 설명변수들을 대상으로 치매발병과 상관성이 있는 요인변수들을 찾기 위해 먼저 단순로지스틱회귀모형을 사용하였다. 다음으로 유의수준 0.05에서 유의한 변수들을 대상으로 선택된 모든 변수들을 가지고 가장 유의하지 않은 각 개별변수를 제거해 나가는 후진 제거법을 이용하여 최종적으로 AIC가 가장 작은 로지스틱회귀모형을 선택하였다. 이러한 후진제거법은 단계적선택법(stepwise selection)이 놓칠 수 있는 최적모형을 선택하는데 도움을 줄 수 있다.

[표 6]은 2015년(연초 기준) 분석대상자를 기준으로 이후 5년 동안 치매발생을 예측하는 다중 로지스틱회귀모형의 모형적합도 통계량을 보여주는 표이다. 선택된 모형의 적합도 지표인 deviance/df는 0.3666(p-값=1.0), Pearson/df는 0.9129(p-값=1.0)에 의한 적합도검정(Goodness of Fit)은 타당함을 보인다. 또한 모형의 설명력을 나타내는 R^2 은 21.99%, AIC=212,499.1이다.

[표 6] 모형적합도검정(Goodness of fit test) 통계량

	Value	DF	Value/DF	p-value
Deviance	24,2418.1	579,538	0.3666	1.0000
Pearson deviance	52,9051.7	579,538	0.9129	1.0000
Likelihood Ratio Chi-square	48.75.4	39		<.0001
Max-rescaled R^2	0.2199			
AIC	212,499.1			

[표 7]은 35개의 변수들 중 단순 로지스틱회귀모형을 적용하여 유의수준 5% 하에서 선택된 변수들 중 후진제거법에 의해 선택된 최적모형에 포함된 28개 변수들에 대한 Type 3 유의성 결과이다. 치매발병 여부에 영향을 주는 요인은 과거질병이력과 여러 건강검진결과 항목들의 28개 요인으로, 치매발병에는 다양한 지표들과 관련성이 있음을 보인다.

체질량지수, 수축기혈압, 이완기혈압의 경우 선형으로 증가하거나 감소하는 형태로 건강상태를 나타내는 것이 아닌, 표준 대비 낮은 구간과 높은 구간으로 벗어나는 경우 건강상태에 영

[표 7] 다중로지스틱회귀모형에 의한 치매발생 유의한 요인

요 인	df	Wald -chisq	p-value
연령	1	23,177.8	<.0001
최근 5년 뇌혈관질환 유병 여부	1	971.6	<.0001
최근 5년 파킨슨병 유병 여부	1	491.5	<.0001
성별	1	438.7	<.0001
최근 5년 입원일수	1	360.3	<.0001
장애중증	2	544.1	<.0001
최근 5년 7일 이상 계속치료 여부	1	237.9	<.0001
최근 5년 입원 여부	1	164.8	<.0001
최근 5년 당뇨병 유병 여부	1	158.4	<.0001
감마지티피(GAMMA_GTP)	1	137.4	<.0001
최근 5년 골절 유병 여부	1	124.5	<.0001
HMG(혈색소, 헤모글로빈)	1	85.8	<.0001
BMI(체질량지수)	1	70.3	<.0001
최근 5년 허혈성심장질환 유병 여부	1	49.3	<.0001
공복혈당(BLDS)	1	48.4	<.0001
BMI(체질량지수)^2	1	42.6	<.0001
최근 5년 수술횟수	1	41.3	<.0001
소득분위	10	356.9	<.0001
수축기혈압(BP_HIGH)	1	36.3	<.0001
HDL콜레스테롤(HDL_CHOLE)	1	25.8	<.0001
혈청GPT ALT(SGPT_ALT)	1	25.7	<.0001
혈청GOT AST(SGOT_AST)	1	19.5	<.0001
총콜레스테롤(TOT_CHOLE)	1	17.7	<.0001
이완기혈압(BP_LWST)	1	16.2	<.0001
DL콜레스테롤(LDL_CHOLE)	1	10.2	0.0014
최근 5년 고혈압 유병 여부	1	7.8	0.0053
최근 5년 수술 여부	1	7.2	0.0072

향을 미친다. 이 점을 보완하기 위하여 체질량지수, 수축기혈압, 이완기혈압 세 개 변수는 각 변수의 제곱(2) 변수를 모형설정시 추가하여 고려하였다. [표 7]을 보면 치매발병에 가장 영향이 큰 요인은 연령이다. 두 번째 요인은 최근 5년 뇌혈관질환 유병 여부이고, 세 번째 요인은 최근 5년 파킨슨병 유병 여부로 나타났다. 다음으로 치매발병 요인은 성별, 최근 5년 입원일수, 장애중증, 최근 5년 7일 이상 계속치료 여부, 최근 5년 입원 여부, 최근 5년 당뇨병 유병 여부, 감마지티피(GAMMA_GTP), 최근 5년 골절 유병 여부, HMG(혈색소, 헤모글로빈), BMI(체질량지수), 최근 5년 허혈성심장질환 유병 여부, 공복혈당(BLDS), BMI(체질량지수)^2,

최근 5년 수술횟수, 소득분위, 수축기혈압(BP_HIGH), HDL콜레스테롤(HDL_CHOLE), 혈청 GPT ALT(SGPT_ALT), 혈청GOT AST(SGOT_AST), 총콜레스테롤(TOT_CHOLE) 순으로 나타났다. 다중로지스틱회귀모형에 기반한 치매발병에 유의한 영향 요인이 총 27개 변수가 선택된 것은 치매와 상관성이 있는 건강지표가 많고 그만큼 치매요인이 복합적인 것으로 사료된다.

[표 8]은 본 연구에서 최근 5년 질병 여부와 건강검진 지표들에 기반 하에 선택된 최적 로지스틱회귀모형의 모수 추정값과 표준오차, p-값, 오즈비를 나타낸 표이다. 노인 치매발병에 가장 큰 요인은 연령으로 나이가 1년 높아질수록 치매발병 오즈는 1.147배 증가함을 보인다. 다음으로 치매발병에 영향을 주로 주는 요인은 최근 5년 동안에 질병의 발생 여부이다.

[표 8] 치매발생 로지스틱회귀모형의 계수 추정치

요 인	구분	df	$\hat{\beta}$	s.e.	p-값	오즈비
상수		1	-7.7110	0.2342	<.0001	-
연령		1	0.1369	0.0009	<.0001	1.147
최근 5년 뇌혈관질환 유병 여부	있음	1	0.4569	0.0147	<.0001	1.580
	없음	0	0.0000	-	-	1.000
최근 5년 파킨슨병 유병 여부	있음	1	1.0145	0.0458	<.0001	2.758
	없음	0	0.0000	-	-	1.000
성별	남성	1	-0.3075	0.0147	<.0001	0.735
	여성	0	0.0000	-	-	1.000
최근 5년 입원일수		1	0.0013	0.0180	<.0001	1.001
장애중증	해당없음	1	-0.2463	0.0180	<.0001	0.782
	중증(1~2등급)	1	0.3451	0.0328	<.0001	1.412
	경증(3~6등급)	0	0.0000	-	-	1.000
최근 5년 7일 이상 계속치료 여부	있음	1	0.2418	0.0157	<.0001	1.274
	없음	0	0.0000	-	-	1.000
최근 5년 입원 여부	있음	1	0.1954	0.0152	<.0001	1.215
	없음	0	0.0000	-	-	1.000
최근 5년 당뇨 유병 여부	있음	1	0.1715	0.0136	<.0001	1.188
	없음	0	0.0000	-	-	1.000
감마지티피(GAMMA_GTP)		1	0.0013	0.0001	<.0001	1.001
최근 5년 골절 유병 여부	있음	1	0.1570	0.0141	<.0001	1.170
	없음	0	0.0000	-	-	1.000
HMG(혈색소, 헤모글로빈)		1	-0.0441	0.0048	<.0001	0.957
BMI(체질량지수)		1	-0.1348	0.0161	<.0001	0.874
최근 5년 허혈성심장질환 유병 여부	있음	1	0.1083	0.0154	<.0001	1.114
	없음	0	0.0000	-	-	1.000

요 인	구분	df	$\hat{\beta}$	s.e.	p-값	오즈비
공복혈당(BLDS)		1	0.0016	0.0002	<.0001	1.002
BMI(체질량지수)^2		1	0.0021	0.0003	<.0001	1.002
최근 5년 수술횟수		1	0.0174	0.0027	<.0001	1.018
소득분위	0분위	1	0.5978	0.0329	<.0001	1.818
	1분위	1	0.1525	0.0240	<.0001	1.165
	2분위	1	0.1158	0.0281	<.0001	1.123
	3분위	1	0.0213	0.0287	0.4573	1.022
	4분위	1	0.0643	0.0300	0.0321	1.066
	5분위	1	0.0786	0.0274	0.0041	1.082
	6분위	1	0.0909	0.0256	0.0004	1.095
	7분위	1	0.0912	0.0238	0.0001	1.096
	8분위	1	0.0616	0.0221	0.0054	1.064
	9분위	1	0.0627	0.0221	0.0017	1.065
	10분위	0	0.0000	-	-	1.000
수축기혈압(BP_HIGH)		1	-0.0032	0.0005	<.0001	0.997
HDL콜레스테롤(HDL_CHOLE)		1	-0.0026	0.0005	<.0001	0.997
혈청GPT ALT(SGPT_ALT)		1	-0.0025	0.0005	<.0001	0.998
혈청GOT AST(SGOT_AST)		1	0.0013	0.0003	<.0001	1.001
총콜레스테롤(TOT_CHOLE)		1	0.0016	0.0004	<.0001	1.002
이완기혈압(BP_LWST)		1	0.0033	0.0008	<.0001	1.003
DL콜레스테롤(LDL_CHOLE)		1	-0.0012	0.0004	0.0014	0.999
최근 5년 고혈압	있음	1	0.0394	0.0141	0.0053	1.040
유병 여부	없음	0	0.0000	-	-	1.000
최근 5년 수술 여부	있음	1	-0.0438	0.0163	0.0072	0.957
	없음	0	0.0000	-	-	1.000

최근 5년 뇌질환이 없는 경우 대비 뇌혈관질환이 있는 경우의 치매발병 오즈는 1.580배 증가함을 보인다. 최근 5년 파킨슨병이 없는 경우 대비 파킨슨병이 있는 경우의 치매발병 오즈는 2.758배 증가한다. 여성 대비 남성의 치매발병 오즈는 0.735배 낮게 나타났다. 장애경증(3~6등급) 대비 장애가 없는 대상자의 치매발병 오즈는 0.782배로 낮아지지만, 장애가 중증(1~2등급)인 경우는 1.412배 높은 것으로 나타났다. 최근 5년 동안 7일 이상 계속 치료가 있는 경우는 그렇지 않은 경우 대비 치매발병 오즈가 1.279배 높다. 최근 5년 동안 입원이 있는 경우는 그렇지 않은 경우 대비 치매발병 오즈가 1.215배 높다. 최근 5년 동안 당뇨병이 있는 경우는 그렇지 않은 경우 대비 치매발병 오즈가 1.188배 높은 경향을 보인다. 최근 5년 골절이 있는 경우는 그렇지 않은 경우 대비 치매발병 오즈가 1.170배 높다. 최근 5년 동안 허혈성심장질환이 있는 경우는 그렇지 않은 경우 대비 치매발병 오즈가 1.115배 높은 경향을 보인다.

다음으로 건강검진 지표들이 노인 치매발병에 영향을 주는 요인들이다. 감마지티피(GAMMA_GTP)가 한 단위 높아질수록 치매발병 오르는 1.001배 높아지고, HMG(혈색소, 헤모글로빈)이 한 단위 낮아질수록 1.045배 높아진다. 공복혈당(BLDS)이 한 단위 높아질수록 치매발병 오르는 1.002배 높아지고, 최근 5년 수술횟수가 한 번 증가할수록 치매발병 오르는 1.018배 높아진다. 총콜레스테롤(TOT_CHOLE)이 한 단위 높아질수록 치매발병 오르는 1.002배 높아지는 경향을 보인다.

소득 또한 노인 치매발병에 영향을 주는 요인으로 나타났다. 소득이 낮을수록 노인 치매발병 가능성은 높음을 보인다. 소득이 가장 높은 소득분위 10분위 대비 소득이 제일 낮은 소득분위 0분위의 치매발병 오르는 1.818배 높고, 소득분위 1분위는 1.165배 높음을 보인다.

[표 9]는 최근 5년 질병 여부와 건강검진 지표들에 기반 하에 로지스틱회귀모형에 의해 향후 5년 동안 치매발생률 추정치와 실제 치매발생률 값을 나타낸 표이다. 모형에 의한 치매발생률과 실제발생률 간에 큰 차이는 크게 보이지 않는다. [표 8]의 노인 치매발생률과 치매발생 모형 예측값은 향후 보험상품 개발과 노인 치매연구에 활용될 수 있을 것으로 사료된다.

[표 9] 노인 치매발생률과 모형 예측값

성별	연령	대상자수	실제		모형예측	
			치매발생률	표준오차	치매발생률	표준오차
남성	55세 ~ 60세 미만	72,442	0.0080	0.0003	0.0091	0.0001
	60세 ~ 65세 미만	70,290	0.0173	0.0005	0.0176	0.0001
	65세 ~ 70세 미만	54,945	0.0389	0.0008	0.0364	0.0001
	70세 ~ 75세 미만	39,241	0.0797	0.0014	0.0740	0.0002
	75세 ~ 80세 미만	24,115	0.1431	0.0023	0.1397	0.0004
	80세 ~ 85세 미만	8,658	0.2267	0.0045	0.2350	0.0010
	85세 이상	2,796	0.2904	0.0086	0.3503	0.1090
여성	55세 ~ 60세 미만	79,945	0.0096	0.0003	0.0125	0.0001
	60세 ~ 65세 미만	79,600	0.0224	0.0005	0.0240	0.0000
	65세 ~ 70세 미만	62,068	0.0516	0.0009	0.0499	0.0000
	70세 ~ 75세 미만	45,988	0.1099	0.0015	0.1019	0.0000
	75세 ~ 80세 미만	29,772	0.1928	0.0023	0.1873	0.0002
	80세 ~ 85세 미만	11,633	0.3002	0.0042	0.3014	0.0011
	85세 이상	4,252	0.3923	0.0075	0.4269	0.0027

V. 결론 및 시사점

본 연구는 국민건강보험공단이 노인코호트DB를 사용하여 개인의 건강검진 데이터와 과거 질병이력 데이터를 이용하여 우리나라 노인들의 치매발병 요인을 찾고 통계적 예측모형을 개발하여 객관적인 치매발병 상대위험도를 추정하고자 하였다.

노인 치매발병에 가장 큰 요인은 연령으로 나이가 증가할수록 치매발병 위험도는 크게 증가함을 보였다. 다음으로 치매발병에 영향을 주로 주는 요인은 최근 5년 동안에 질병의 발생 여부이다. 최근 5년 동안에 뇌질환, 파킨슨병 유병여부, 7일 이상 계속 치료, 입원, 당뇨병, 골절 유병, 허혈성심장질환 유병, 수술, 고혈압이 있을수록 노인 치매발병 위험이 높음을 보였다.

다음으로 건강검진 지표들이 노인 치매발병에 영향을 주는 요인들이다. 감마지티피, HMG(혈색소, 헤모글로빈), 공복혈당(BLDS), 총콜레스테롤, 수축기혈압, 이완기혈압, 혈청GPT(ALT), 혈청GOT(AST), 요단백 등은 치매발병에 영향을 주는 중요한 지표임을 보였다. 이러한 건강검진 결과는 치매발병 위험을 예측할 수 있는 중요한 지표가 될 수 있음을 나타낸 것이다.

성별 또한 노인 치매발병에 큰 영향을 주는 요인으로, 여성이 남성보다 치매발병 위험이 높게 나타났다. 노인 치매발병에 또 다른 주요 영향 요인은 장애중증 요인으로 장애가 있고, 장애등급이 높을수록 노인 치매발병 위험은 높은 것으로 나타났다. 소득 또한 노인 치매발병 위험에 영향을 주는 요인으로 나타났다. 소득이 낮을수록 노인 치매발병 위험은 높음을 보인다.

본 연구의 중요한 의미와 시사점은 다음과 같다.

첫째, 치매는 무엇보다도 치방이 아닌 예방이 중요한 질병으로 익히 알려져 있다. 본 연구에서 과거질병이력과 건강검진이력이 치매발병 위험에 절대적 전조 역할을 하고 있음을 국민건강보험공단의 노인코호트DB를 이용하여 실증적으로 검증하였다. 이는 대한민국 국민들이 치매를 방지하고 예방하기 위해 어떠한 건강검진 지표들을 관리하고 어떠한 질병들을 예방해야 하는지를 제시하고 있다.

둘째, 본 연구에서는 과거에는 사용되지 않은 건강검진데이터와 과거 질병이력에 기반 하에 개발된 치매발병 예측모델은 개인의 건강 데이터를 종합적으로 분석하여 치매의 조기 발견과 예방에 중요한 도구가 될 수 있다. 기존 치매 발병 위험을 예측하는 연구는 주로 고비용의 정

밀한 데이터를 활용한 방식에 의존했으나, 본 연구에서는 보다 실용적이고 접근이 용이한 건강검진 정보와 과거 질병 이력을 바탕으로 치매 발병 예측모형을 개발하였다. 이러한 접근법은 경제적 부담을 줄이면서도 일상적으로 활용 가능한 데이터를 통해 치매 위험을 평가할 수 있어, 조기 발견과 예방에 효과적인 도구로 기능할 수 있음을 보여주었다.

셋째, 현재 국내 보험사에서는 치매관련 상품을 개발하는데 있어 아직까지 판매 기간이 짧아 자체적으로 장기적인 치매의 진전 및 추이에 대한 객관적인 통계 집적이 이루어지지 않고 있다. 향후 보험사는 본 연구에서 제시한 국민건강보험 노인코호트DB를 근거한 치매발병 통계적 모형에 참고하여 과거 병력 및 건강상태에 따라 치매발병 위험을 치매보험에 적용할 수 있는 중요한 기초자료로 사용될 수 있을 것으로 사료된다. 본 연구에서 제시한 국민건강보험 노인코호트DB를 기반으로 한 실제 치매 발병 위험은 기존 보험사에서 사용하는 치매 위험률에 할증·할인 요소로 적용이 가능하여 다양한 보험계약자 집단에 보다 현실적이고 정확한 치매보험 상품을 개발할 수 있을 것이다.

다섯째, 현재 국내 보험사들은 치매담보 보험상품 개발에 일본인들을 대상으로 생성된 치매요율을 이용하고 있다. 향후 국내 보험사들은 본 연구에서 개발된 치매요율을 사용함으로써 다양한 치매관련 보험상품 개발에 사용될 수 있을 것으로 기대된다.

여섯째, 본 연구에서 개발된 치매발병률 예측모형을 통해 치매예방연구에 새로운 방향을 제시하고 노인들의 공중보건향상에 기여할 수 있을 것으로 사료된다.

다만, 본 연구는 치매의 진단을 치매 상병코드만을 사용하여 치매 여부를 판별하였으나, 실제 치매는 치매 상병코드뿐만 아니라 치매임상평가척도(Clinical Dementia Rating, CDR)라는 치매 진단 방법을 적용하여 경증, 중등도, 중증으로 나뉘어 평가되고 있다. 그러나 본 연구에서 사용된 코호트 데이터에는 CDR 정보가 포함되어 있지 않아, 진단의 정확성과 치매의 정도를 세부적으로 반영하는 데 한계가 있으며, 이로 인해 실제 치매 발생률 산출에도 제한이 따른다. 향후 연구에서는 CDR 점수를 포함한 다각적인 진단 방법을 활용하여 보다 정밀한 분석을 통해 치매의 유병률 및 예후를 평가하는 것이 필요할 것이다.

결론적으로 본 연구에서 제시된 치매발병 위험의 분석 결과는 치매 리스크에 노출될 수 있는 미래 치매환자들과 그 가족들의 경제적 리스크를 준비할 수 있는 보험상품 개발에 기여할 수 있을 것이다. 또한 국민건강보험 노인코호트DB를 사용하여 실제 노인들의 과거 질병이력과 건강검진 지표들을 통해 치매발병률 예측 모델의 개발을 통해 치매 예방 연구에 새로운 방

향을 제시하고 노인들의 공중보건 향상에 기여하고자 한다.

그러나 본 연구는 다음과 같은 연구의 한계점이 존재함을 제시하고자 한다. 첫째, 모든 국민 전체를 대상으로 한게 아니라 건강검진 지표들을 이용하기 위해 건강검진을 받은 대상자들을 연구대상으로 한정했다는 한계점이 존재한다. 둘째로 본 연구에서는 주상병과 부상병의 치매 진료이력이 있는 대상자들을 치매발병으로 명명하였지만 급여항목인 치매약제성분의 정보를 사용하지 않아 실제 치매 발병보다 과도한 인원이 정의 될 수 있다(Huh, 2024). 셋째, 본 연구는 기준년도를 정하여 향후 5년 특정 기간에 치매 발병을 조사하였지만 Cox-비례위험모형을 이용하면 치매요인과 특정 나이 이후의 치매발병 기간을 구할 수 있을 것이다. 마지막으로, 본 연구는 국민건강보험공단 자료를 활용한 연구의 제한점으로 진단의 정확성, 잠재적인 교란 요인, 비급여 항목의 부재, 선택적 편향 등에 대한 한계점을 가질 수 있음을 제시하고자 한다.

참 고 문 헌

- 권혁성, 2012, 노인성 질병에 대한 민영 보험의 수요추정과 보험수리 모형의 설계에 관한 연구, 보험학회지, 93, 107-129.
- 김민국, 2019, 국내외 치매보험 시장 현황과 우체국보험의 시사점, 우정정보, 2019권 제3호, 63-86.
- 이태인 · 오하영, 2021, 이기종 머신러닝 모델 기반 치매예측 모델, 한국정보통신학회논문지, 제25권 제12호, 172-1738.
- 한수정 · 백남중, 2015, 치매의 약물치료, Geriatric Rehabilitation. 제8권 제1호, 19-23.
- 보건복지부, 2016년 전국 치매역학 조사
- 국립중앙의료원, 2024, 대한민국 치매현황 2023, 중앙치매센터.
- 조영은 · 우종필 · 김중안, 2024, 머신러닝 기법을 활용한 치매 예측 모델과 상업적 활용 전략 : 웨어러블 기기의수면 및 활동 데이터를 기반으로, 경영정보학연구, 제26권 제2호, 137-53.
- 질병관리청, 2024, <https://health.kdca.go.kr>
- Agresti. A., 2013, Categorical data analysis (3rd ed.). New York.: John Wiley & Sons.
- Benito-León, J., Louis, E. D., Villarejo-Galende, A, Romero, J.P., and Bermejo-Pareja, F., 2014, Long sleep duration in elders without dementia increases risk of dementia mortality (NEDICES). Neurology, 83(17), 1530-7.
- Henry, A., Katsoulis, M., Masi, S., Fatemifar, G., Denaxas, S., Acosta, D., Garfield, V., and Dale, C.E., 2019, The relationship between sleep duration, cognition and dementia: a Mendelian randomization study. Int J Epidemiol, 48(3), 849-860.
- Huh, H., Kim, M., Jung S., Cho, J.M., Kim, S.G., Park, S., Lee, S., Kang, E., Kim, Y., Kim, D.K., Joo, K.W., Han, K., and Cho, S., 2024. Menopausal hormone therapy and risk for dementia in women with CKD: A nationwide

- observational cohort study. *Nephrology* (Carlton). 29(3), 126-134.
- Jorm A. F., 1991, Cross-national comparisons of the occurrence of Alzheimer's and vascular dementias. *Eur Arch Psychiatry Clin Neurosci*. 240(4-5), :218-22.
- Leng, Y., Musiek, E.S., Hu, K., Cappuccio, F.P., and Yaffe, K.. 209, Association between circadian rhythms and neurodegenerative diseases. *Lancet Neurol*, 18(3):307-318.
- Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., and Buckner, R.L., 2007, Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci*. 19(9), 1498-507
- Navrtilov, M., Jarkovsk, J., Cešková, E., Leonard, B., and Sobotka, L., 2007, Alzheimer disease: malnutrition and nutritional support. *Clinical and experimental pharmacology and physiology*, 34(s1), S11-S13.

I금융감독연구 제12권 제1호 2025. 4.

Development of a Dementia Onset Prediction Model Based on Past Medical History and Health Check-Up Indicators

– Using the National Health Insurance Service's Elderly Cohort DB

Heuju Chun*, Taekyo Leen**

Abstract

This study aimed to identify the factors contributing to dementia onset among the elderly in South Korea by utilizing individual health check-up data and past medical history data from the National Health Insurance Service's Elderly Cohort Database. Additionally, it sought to develop a statistical prediction model to estimate the objective incidence rate of dementia. The analysis revealed that age is the most significant factor influencing dementia onset, with the likelihood of developing dementia increasing considerably as age rises. The second major factor affecting dementia onset is the occurrence of diseases in the past five years. Higher incidence of conditions such as brain diseases, Parkinson's disease, hospitalizations and diabetes was associated with a higher rate of dementia among the elderly. Health check-up indicators were also presented as significant factors influencing dementia onset. The estimated dementia incidence rates presented in this study could be utilized in developing insurance products to prepare future dementia patients and their families for the economic risks associated with dementia.

Key words: Health check-up, past medical histories, dementia onset, logistic regression
JEL Classifications:: J32

* First author, Associate Professor, Department of Statistics and Information, Dongduk Women's university(hjchun@dongduk.ac.kr)

** Corresponding author, Ph.D, Hanyang University(tkleen@naver.com)