

당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

박영민 · 이상아 · 신동교 · 임현선 · 김보연 · 배세진

h·well
국민건강보험



국민건강보험

National Health
Insurance Service

일산병원

Ilisan Hospital

공동연구보고서

NHIMC-2021-PR-022

당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

박영민 · 이상아 · 신동교 · 임현선 · 김보연 · 배세진



[저 자]

책임 연구자:	국민건강보험 일산병원 가정의학과	박영민
	국민건강보험공단	이상아
공동 연구원:	국민건강보험 일산병원 연구소 연구분석부	임현선
	국민건강보험 일산병원 의무기록부	신동교
	국민건강보험 일산병원 의무기록부	김보연
	국민건강보험공단	배세진

연구관리번호

IRB 번호

NHIS-2022-1-008

NHIMC 2021-07-001

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

머리말

4차 산업혁명 시대에 첨단 의학의 기반은 곧 데이터 분석 능력이며 임상 데이터를 체계적으로 축적하고 활용할 수 있는 시스템이 곧 경쟁력이 된다는 말이 있다. 이렇듯 4차 산업혁명 시대의 키워드로 꼽히는 임상 빅데이터는 의료계로도 산업계에서도 매우 중요한 이슈 중의 하나이다. 실제로 최근 임상 빅데이터에 대한 관심이 높아지면서 대형병원들을 중심으로 임상데이터웨어하우스(Clinical Datawarehouse)를 구축해 연구 기반을 쌓으려는 움직임이 활발해지고 있는 상황이다.

또한 앞서 지난 2018년 정부는 데이터 경제 활성화 규제혁신 정책으로 바이오헬스 빅데이터 구축을 위한 산업단을 추진을 밝히고 지금까지 그 사업을 시행해오고 있다. 이는 병원들이 보유한 의료 데이터를 공통데이터모델(CDM)로 표준화하고 데이터를 분석하기 위한 소프트웨어 개발과 플랫폼 구축을 목표로 하고 있다.

이렇듯 의료빅데이터 활용을 위한 연구정보시스템의 구축과 그 활용을 앞으로 국민건강보험 일산병원은 물론 많은 병원들이 나아가야 할 중요한 방향 중 하나이다. 이번 연구를 통해서 일산병원의 두 연구정보시스템의 현황을 알아보고 그 나아갈 바를 고민해보고자 한다.

끝으로 이 보고서의 내용은 연구진의 개인적 의견이며 국민건강보험 일산병원의 공식적 견해가 아님을 밝혀둔다.

2022년 4월

국민건강보험공단 이사장

강도태

국민건강보험 일산병원장

김성우

국민건강보험공단 건강보험연구원장

이해종

국민건강보험 일산병원 연구소장

이천균

목차

요약	1
제1장 서론	11
제1절 연구 배경 및 필요성	13
제2절 연구 목적	16
제2장 이론적 고찰	17
제1절 보건의료 정보화 실태조사 현황	19
제2절 국민건강보험 일산병원 임상데이터웨어하우스	21
제3절 국민건강보험 일산병원 공통데이터모델	24
제4절 당뇨병환자에서 저밀도콜레스테롤 조절과 심혈관질환	27
제3장 연구 내용 및 방법	33
제1절 연구 내용	35
제2절 연구 방법	39
제4장 연구 결과	45
제1절 국민건강보험 일산병원 CDM 현황	47
제2절 국민건강보험 일산병원 CDW 현황	52
제3절 당뇨병환자에서 Statin제제 사용에 따른 LDL-C와 심혈관질환 위험 및 사망위험 분석	56
제5장 고찰 및 제언	63
제1절 결론 및 고찰	65
제2절 제언	68
참고문헌	73
부록	79

표목차

<표 2-1> 국민건강보험 일산병원 버전별 CDM 데이터베이스 특징	27
<표 2-2> ATP III 가이드 라인	30
<표 3-1> CDM 테이블(도메인)별 설명	36
<표 3-2> 대상자 관련 변수 및 관련내용	40
<표 3-3> 혼란 변수 및 관련내용	41
<표 4-1> 국민건강보험 일산병원 CDM 5.3.0_02버전 매핑률	47
<표 4-2> 아주대병원 CDM 5.3.0 버전 매핑률	48
<표 4-3> 국민건강보험 일산병원 CDM 5.3.1_weekly 버전 매핑률	49
<표 4-4> 국민건강보험 일산병원 CDM 5.3.0_02버전 진단검사의학적 변수의 매핑률	50
<표 4-5> NHIMC 5.3.0_02버전 데이터베이스내 흡연관련 Concept 현황	51
<표 4-6> 구DW, CDW 및 iDEA에서 산출된 대상자수 비교	52
<표 4-7> KCD7 E119 코드에 대한 세부사항 예	54
<표 4-8> CDM데이터베이스를 활용한 당뇨병환자에서 Statin사용여부에 따른 심혈은 질환 및 사망위험(PS 매칭전후)	56
<표 4-9> 연구대상자 기본특성	58
<표 4-10> Statin 사용 18개월후 콜레스테롤 수치 변화	59
<표 4-11> DW기반 공단연계자료에서 Statin사용군과 비사용군간의 심혈관질환 발생 및 사망위험	59
부록표 1. PCI 및 CABG의 CDM내 Concept ID	81
부록표 2. Brain CT 및 MRI의 CDM내 Concept ID	82

그림목차

[그림 1-1] 임상연구의 개요	13
[그림 2-1] CDW의 ETL(데이터추출) 주기	20
[그림 2-2] CDW의 제공데이터 종류	20
[그림 2-3] CDM 변환 및 표준용어 매핑	21
[그림 2-4] 데이터 웨어하우스의 개요도	22
[그림 2-5] 국민건강보험 일산병원 DW와 CDW의 모식도	23
[그림 2-6] 분산형 연구망	25
[그림 2-7] 2008년 미국 당뇨병환자에서의 사망원인	27
[그림 2-8] 당뇨병환자 및 비당뇨병환자에서 심근경색 사망에 대한 위험에 대한 생존분석 곡선	28
[그림 2-9] Atrovastatin 사용여부에 따른 심혈관질환 및 사망위험	29
[그림 2-10] 당뇨병환자에서 고콜레스테롤혈증 유병률과 조절률	31
[그림 3-1] OMOP-CDM 구조 개요도	35
[그림 3-2] 연구의 틀	39
[그림 4-1] 국민건강보험 일산병원 CDM 5.3.0_02버전 인원당 각 테이블(도메인)의 Concept ID수	49
[그림 4-2] iDEA 내의 조건그룹 실행오류 예	54
[그림 4-3] CDW 데이터베이스내 흡연 및 체질량지수 데이터 현황	55
[그림 4-4] 연구대상자 흐름도	57
[그림 4-5] DW기반 공단연계자료에서 Statin사용군과 비사용군간의 LDL-C에 따른 심혈관질환 발생 및 사망위험	61

요약



요약

1. 연구 배경 및 목적

실험연구와 전향적 관찰연구는 그 실행에 있어 다양한 제약들이 따르는 가운데, 기존의 존재하는 임상의료정보를 바탕으로 한 후향적 연구에 대한 관심이 많아지고 있다. 이에 여러 기관의 자료를 통해서 어떤 연구결과가 동일하게 반복되는 재현성을 통해서 타당도를 확보하면서도 개인정보보호의 문제, 자료간의 이질성 등으로 인해서 현실적으로 어려움을 해결하기 위한 방법으로 공통데이터표준을 통해 데이터의 구조를 통일하고 연구자의 분석환경으로 데이터를 추출하지 않고 연구자가 만든 분석프로세스만을 다른 기관의 데이터가 있는 곳으로 전달하는 방식으로 공통데이터모델(common data model, CDM)에 개발되었다. 또한 전자의무기록 시스템으로 부터의 정보가 의사결정을 지원하기 위해서는 시간의 선후관계를 가지고, 정보의 주제별로 분류, 통합된 데이터 저장소가 바로 데이터웨어하우스(data warehouse, DW)이며, 특히 최근 병원에서 환자 진료중에 발생하는 임상자료와 분석도구를 의료진에게 제공하여 연구를 지원하는 시스템을 임상데이터웨어하우스(clinical data warehouse, CDW)라고 한다.

국민건강보험 일산병원은 최근에 미래를 위한 정보화 사업에 힘입어 CDM과 CDW를 모두 구축하였으며, 공단자료와 연계하여 연구에 활용할 수 있다는 연구에 있어 중요한 이점을 가지고 있다. 하지만 아직까지 구축 후에 연구를 통한 적절한 검증이 이루어지지 않은 상황이다.

반면에 당뇨병은 건강문제에 있어서 가장 중요한 만성질환의 하나로 당뇨병환자에게 심혈관질환 및 사망 위험을 감소시키는 방법으로 여러 진료지침을 통해서 LDL-Cholesterol(LDL-C)를 낮추는 것이 추천된다. 실제 임상현장에서의 당뇨병환자에서 statin 처방여부에 따른 심혈관질환의 위험과 사망위험을 살펴볼 필요가 있다.

이에 본 연구에서는 일산병원 CDM과 공단자료를 연계한 DW 자료 활용하여 일산병원 당뇨병환자 statin 제제의 복용여부와 이에 따른 심혈관질환의 발생위험 및 사망위험을 확인해 보고자 한다. 또한, 이 연구과정과 각 변수들이 연구 과정을 통해서 활용되는

CDM, CDW 데이터베이스의 현황과 연구과정에서의 문제점들을 통해 향후 연구자들이 해당 자료를 활용할 때 고려해야할 사항들에 대해서도 확인하여 추후 CDM CDW를 활용한 연구자들의 연구진행에 도움이 되고자 한다.

2. 연구 결과

국민건강보험 일산병원 CDM 5.3.0_02 버전 데이터베이스는 검사결과의 Measurement 도메인의 73.9%를 제외하고 대부분의 도메인에서 레코드의 90%이상이 매핑되어 있었다.

<표 1> 국민건강보험 일산병원 CDM 5.3.0_02버전 매핑률

(2021. 12. 31. 현재)

도메인	원본 레코드수	매핑된 레코드수	매핑 레코드 비율 (%)	원본 코드수	매핑된 코드수	매핑코드 비율 (%)
Person	1,367,171	1,367,171	100.0%			
Condition_occurrence (진단)	26,432,955	26,307,899	99.5%	15,592	14,328	91.9%
Drug_exposure (약물)	83,237,465	83,027,507	99.7%	5,143	4,932	95.9%
Measurement (검사)	290,183,181	278,217,168	95.9%	2,984	2,204	73.9%
Procedure_occurrence (시술처치)	103,313,825	86,622,017	83.8%	16,831	14,306	85.0%
Visit_occurrence (진료방문)	16,577,348	16,576,618	100.0%	4	4	100.0%
Observation_period (추적기간)	1,367,171	1,367,171	100.0%			
Specimen (검체)	22,922,827	22,706,054	99.1%	403	357	88.6%
Observation (조사정보)	7,605,094	5,963,700	78.4%	308,347	308,284	100.0%
Death	21,130	21,130	100.0%			

4 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

이번 연구에 사용되었던 진단검사의학적 변수인 LDL-C, Triglyceride, Hba1c, Albumin-Creatinine ratio, GFR 등의 변수들도 모두 100%의 매핑률을 부여주었다. 하지만, 새롭게 구축된 5.3.1 weekly 버전의 경우 같은 대상자 조건에 대하여 5.3.0_02버전에 비해 인원수가 적게 산출되어 낮은 매핑률로 인해 아직까지 연구에의 활용은 제한적일 것으로 판단된다.

<표 2> 구DW, CDW 및 iDEA에서 산출된 대상자수 비교

	Definition	factor	KCD7 code	추가정보	CDW 인원수	i-DEA 인원수	DW(구) 인원수
1	Type 2 DM	diabetes	E11-14	CDW상 첫 진단을 구하기 힘들			
2	2002-2015	진단일자			39,414	39,414	39,414
3	Age 30-79	진단당시 나이		진단시 나이를 case design에서 구하게 되면 실제 초진일 누락 가능성			
		anti-diabetic drugs		첫 진단일 이후 10년 이내 (3650일) 진단받은 내역	33,669	33,669	33,669
4	기저 콜레스테롤 수치	LDL-cholesterol > 70mg/dl		당뇨병 진단전 6개월 내에 LDL-C ≤ 70mg/dl 또는 당뇨병 진단전 6개월 내에 500 mg/dl ≤ TG ≤ 2000 mg/dl 인 환자 제외	1,767	1,767	1,763
5		Triglyceride < 500mg/dl			391	391	391
6	과거력자 배제	myocardial infarction	I21, I22	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역	4,433	4,433	4,433
7		angina pectoris	I20	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역			
8		cerebral infarction	I63	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역			
9		Transient ischemic attack	G458, G459	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역			
10		placement of coronary stent		당뇨병 첫 진단 이전 10년까지의 기간 중 처방받은내역			
11	Aortocoronary artery bypass graft		당뇨병 첫 진단 이전 10년까지의 기간 중 처방받은 내역	792	792	792	
12	과거 이상지혈증 약물복용자 배제	all Statin		당뇨병 첫 진단 이전 2년까지의 기간 중 처방받은 내역	6,816	6,816	6,816
		fibrate		당뇨병 첫 진단 이전 2년까지의 기간 중 처방받은 내역			
13	고혈당자 배제	Hba1c < 12%		모든 시간	2,624	2,624	2,624

국민건강보험 일산병원 CDW에 대해서는 DW로부터 정리된 대상자조건에 대해서 Sybase, CDW 데이터 베이스로부터 MSSQL프로그램을 통해서 인원수를 산출하였다. 마지막으로 CDW의 웹인터페이스에 해당되는 iDEA를 이용하여 각 조건에 대한 인원수를 산출하여 각각을 비교하였다. 각조건들에 대해서 동일한 인원수가 산출됨을 확인하였다. 현재 새롭게 구축된 CDW 데이터베이스에는 구DW 시스템의 데이터가 상당수준에서 잘 이관되었음을 확인하였다. 또한, CDW의 웹인터페이스인 iDEA를 통해서 CDW의 대상자 산출도 잘되고 있음을 확인할 수 있었다.

DW를 기반으로 한 공단연계자료 분석에서 시행한 분석에서 2004년부터 2015년까지 국민건강보험 일산병원에서 진단받은 30-79세의 당뇨병환자 22,613명에 대해서 2002년부터 2020년까지의 공단자료를 연계하였다. 그중 과거 심혈관질환 등의 과거력, statin제제를 복용력 등을 제외하고 17,407명에 대해서 최종적으로 분석을 시행하였다. 당뇨병 진단 2년 이내 1년 이상 연속적으로 Statin을 먹은 1,809명, Statin제제를 당뇨병 진단후 2년동안 처방이 없었던 비사용군 1,414명을 대상으로 분석하였다.

<표 3> Statin 사용 18개월후 콜레스테롤 수치 변화

	N	LDL-C		Control rate for <100 mg/dl	% Change of LDL-C from Base	Non-HDL		Control rate for <130 mg/dl	% Change of non-HDL-C from Base
		base	>18m			Base	>18m		
Non-Statin	1414	105.5	99.3	52%	-4.3%	132	124.6	58%	-3.2
Statin	1809	134.4	94.3	61%	-27.9%	167.2	120.3	63%	-28.3

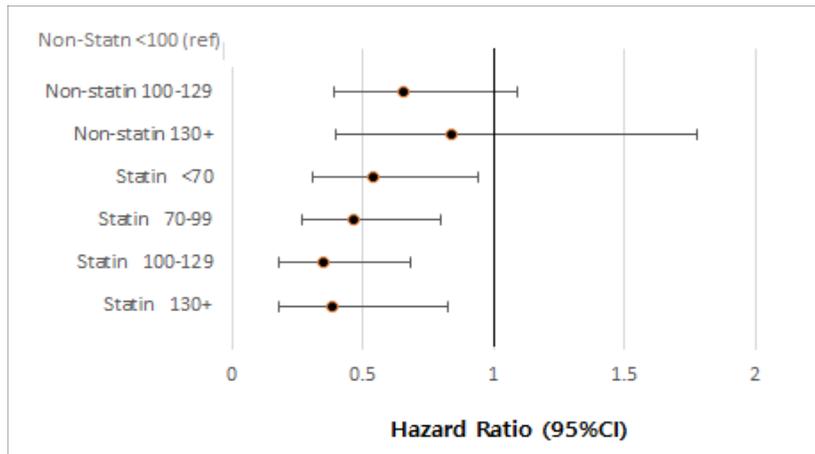
기저 LDL-C 수치가 Statin사용군은 평균 134mg/dl였고 비사용군은 106mg/dl 였다. Statin 사용군에서 처음 사용으로부터 18개월후 LDL-C 수치는 61%, Statin 비사용군의 52%에서 100mg/dl미만으로 조절되었다. 또한, stain 사용군에서 기저치에 비해서 LDL-C가 27.9%정도 감소한 수치를 보였다.

<표 4> DW기반 공단연계자료에서 Statin사용군과 비사용군군의 심혈관질환 발생 및 사망위험

	non-Statin users(N=1414)			Statin users(=1809)			
	Event n(%)	Person years	IRate per 1000py	Event n(%)	person year	IRate per 1000py	Hazard Ratio(95%CI)*
primary Outcome							
Acute Myocardial infarction	7(0.3)	13207	0.53	39(2.1)	14655	2.66	0.86 (0.358-2.082)
Coronary revascularization	7(0.3)	13210	0.53	69(3.8)	14221	4.85	0.75 (0.339-1.654)
Ischemic Stroke	1(0.1)	13303	0.07	11(0.6)	15094	0.72	0
hemorrhagic Stroke	2(0.1)	13305	0.15	2(0.1)	15129	0.13	0
Secondary outcome							
All cause Death	259 (18.3)	10316	25.1	176(9.6)	14731	11.94	0.52 (0.419-0.640)

연령 성별 보정을 했을때 스타틴군에서 급성심근경색, 관상동맥 재관류술 발생위험이 CDM 데이터베이스 결과와는 반대로 statin 사용군에서 발생위험이 각각 0.86배, 0.75배로 감소하였으나, 통계적으로 유의하지는 않았다. 모든 원인의 사망에 대해서는 비교한 결과, statin 사용군에서 0.52배로, 비사용군에 비해서 48%의 사망위험이 감소하는 것으로 나타났다.

마지막으로 statin 비사용군에서 LDL-C < 100 mg/dl 인 그룹을 참고군으로 하여 statin 사용군에서 statin사용후 18개월 이후의 수치로 LDL-C 수치별 그룹에서의 심혈관질환 발생위험 및 사망위험을 분석하였다. LDL-C 100mg/dL 미만인 statin 비사용군과 비교할 때 모든 원인의 사망에 대해서 LDL-C 100mg/dl 미만인 statin 비사용군에 비해 statin을 사용군의 모든 범위의 LDL-C 그룹에서 사망이 유의미하게 감소하였다.



[그림 1] DW기반 공단연계자료에서 Statin사용군과 비사용군간의 LDL-C에 따른 사망위험

3. 결론 및 제언

Statin 비사용군에서 보여준 특징 중 하나는 기저치가 LDL-C 105.5mg/dL로 사용군의 134mg/dL에 비해서 확연히 낮았다. LDL-C 수치가 많이 높지 않은 군이 statin 사용되지 않는 경향이 있음을 보여준다. LDL-C 100mg/dl미만인 statin 비사용군에 비해 모든 LDL-C 수치의 statin 사용군의 사망률이 유의미하게 감소하였다. 100mg/dl 정도로 LDL-C의 수치를 유지하는 경우에도 보다 적극적으로 statin제제를 당뇨병환자에서 사용하는 것이 모든 원인의 사망을 감소시키는 등의 의미 있는 당뇨병환자 관리 전략이 될 수 있음을 보여준다.

CDM 데이터베이스 활용을 위한 다양한 형태 디자인의 연구시도 및 연구 디자인 적용을 통한 데이터베이스 오류 확인 및 수정이 필요하다. CDM 데이터베이스는 분산형 연구망을 통한 여러 병원과의 협력연구가 가능하다는 특징을 가지고 있지만, 근본적으로는 국가단위의 결과값을 알기 어려운 한 병원 단위의 자료임을 확인하고 주로 약물의 안전성 및 부작용 등 노출로부터 길지 않은 시간내에 결과값이 나오는 연구가 아직은 효과적일 수 있다.

연구자들이 보다 지속적으로 CDM 데이터베이스에 관심을 가지고 다양한 연구를 시도해보기 위해서는 보다 꾸준한 지원이 필요하다.

현 CDW 시스템이 구축되기까지 CDW 구축단을 중심으로 많은 노력을 기울여져 왔다. 연구자들이 실제로 연구과제를 현 CDW와 iDEA를 통해서 적용해보고 그 과정에

서 발견되는 오류와 개선점을 피드백을 통해서 전달하고 이를 다시 시스템에 적용하는 선순환적인 구조가 필요하다. CDW가 연구자친화적인 구조를 가지고 있고 iDEA를 통해서 구DW보다 손쉽게 대상자를 산출하고 그 특성들을 확인할 수 있다고 해도 연구가 잘 진행되기 위해서는 계속해서 관련부서의 지원이 필수적이다.

CDW 데이터 베이스가 일개병원의 데이터라는 한계점을 가지고 있으나, 일개병원안에서도 충분한 대상자를 확보할 만한 질병군 또는 이를 바탕으로 레지스트리 연구가 좋은 연구방법이 될수 있다. 나아가 본 연구에서 진행했던 것처럼 CDW를 바탕으로 한 자료에 일산병원의 강점으로 공단자료가 연계되어 연구에 활용된다면 단일 기관 자료로서의 한계를 좀더 보완할 만한 연구가 진행될 수 있으리라 본다.

제 1 장

서론

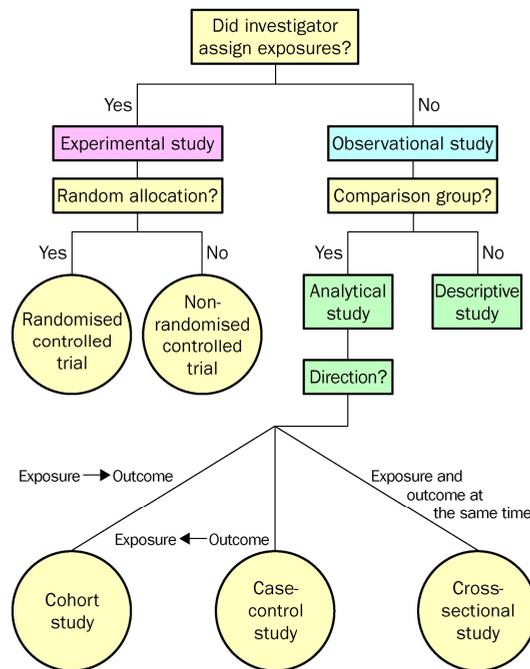
제1절 연구 배경 및 필요성	13
제2절 연구 목적	16

제1장 서론

제1절 연구 배경 및 필요성

1. 임상연구와 임상 연구정보 시스템¹⁾

임상연구는 다양한 방법론으로 시행될 수 있는데 그중에서도 실험연구와 관찰연구 혹은 비중재연구로 분류될 수 있다. 실험연구는 연구자가 연구를 시행함에 있어 변수들을 통제함으로써 여러 가지의 뻘뻘림과 교란요인들을 최소화할 수 있는 연구방법이다. 하지만, 연구수행의 난이도, 비용 문제, 윤리문제 등의 여러 가지 난관들이 존재하여 주로 신약이나 신기술을 중심으로 한 임상 연구에 국한하고 있다. 이에 반면에 관찰연구는 [그림 1-1]에서 보는 것처럼 실험연구와 같이 특정 노출에 대해서 중재 없이 연구를 진행한다¹⁾.



[그림 1-1] 임상연구의 개요¹⁾

1) 출처: 0.1016/S0140-6736(02)07283-5

관찰연구를 다양한 형태로 분류할 수 있다. 일단 비교군·대조군이 있으면 분석연구(analytical study)라 하고, 없는 경우 기술연구(descriptive study)라 하며, 그중에서 분석연구는 노출과 결과의 시간적 선후관계에 따라 코호트연구(cohort study), 환자-대조군 연구(case-control study) 및 단면연구(cross-sectional study)로 나뉜다. 특히 관찰연구 중에서 연구에 따라서는 정해진 집단에서 전향적으로 관찰연구를 시행하는 경우도 있지만, 자료 수집을 위해 비용과 시간면에서 상당한 제약이 뒤따른다. 이에 요즘에는 이미 수집된 이차자료를 이용하는 후향적 연구가 특히 환자가 받은 의료정보를 바탕으로 많이 이루어지고 있다. 이러한 후향적 연구는 자료수집에 할애되는 시간과 비용을 크게 줄일 수 있을 뿐만 아니라, 실험연구에서 시행되는 변수의 통제나 중재 때문에 수행하기 오히려 어려운 다양한 상황에 대한 분석이 가능하다. 그럼에도 관찰연구의 경우 최근까지도 직접적인 의무기록 검토를 수행해야 하는 번거로움 때문에 대규모의 연구대상자를 대상으로 연구하기에 어려움이 있었다.

하지만 최근 전자건강기록이 보급화되고 임상연구정보시스템이 구축되면서 이러한 연구정보시스템을 바탕으로 한 임상자료 활용한 연구의 기초가 세워지고 있다. 연구정보시스템은 임상진료를 위한 임상정보시스템에 대응하여 연구 혹은 2차 활용 지원을 위한 정보시스템을 의미하는 것으로, 임상시험관리시스템(Clinical Trial Management System, CTMS), 임상데이터웨어하우스(Clinical Data Warehouse, CDW), 유전체정보 분석시스템 등이 포함된다.

2. 공통데이터모델(common data model, CDM)

한 의료기관이 가진 임상의료정보는 의료보험의 청구데이터에 비해서 다양한 검사결과 및 처치 등의 정보를 활용할 수 있다는 장점을 지니고 있다. 이에 많은 연구자들이 여러 기관의 임상의료정보를 익명화하고 이를 통합적으로 연구할 수 있는 방법에 대해서 많은 고민을 해왔다. 여러 기관의 자료를 통해서 어떤 연구결과가 동일하게 반복되는 재현성을 통해서 타당도를 확보할 수 있기 때문이다. 하지만 이렇게 여러 기관의 자료를 확보하는 것은 개인정보보호의 문제, 자료간의 이질성 등으로 인해서 현실적으로 어려움을 가지고 있다.

이러한 현실적인 어려움을 타개하기 위한 방법으로 공통데이터표준을 통해 데이터의 구조를 통일하고 이러한 공통 데이터 표준은 연구자의 분석환경으로 데이터를 추출하지 않고 연구자가 만든 분석프로세스만을 다른 기관의 데이터가 있는 곳으로 전달하는 방식으로 작동이 가능하도록 하였다. 이에 각 기간에서의 데이터 추출 단계를 생략하고 데이

터에 대한 표준화된 분석을 실행함으로써 개인 정보 문제를 줄여줄 수 있다. 이러한 데이터의 표준이 공통데이터모델(common data model)에 의해 제공된다.

3. 임상데이터웨어하우스

원래의 전자의무기록 시스템은 임상연구에 최적화된 시스템은 아니다. 환자의 진료 및 진료와 관련된 행정업무를 처리하는 것을 지원하는 것이 주된 목적이다. 따라서 임상 연구를 목적으로 하여 전자의무기록 정보 자체를 활용하기에는 기능적으로나 접근성 면에서 떨어질 수밖에 없다. 따라서 임상정보를 연구에 활용하기 위해서는 전자의무기록의 복사본에 해당되는 정리된 데이터베이스와 분석시스템이 필요하다. 또한 의사결정을 지원하기 위해서는 시간의 선후관계를 가지고, 정보의 주제별로 분류, 통합된 데이터 저장소가 바로 데이터웨어하우스(data warehouse, DW)이다. 특히 최근 병원에서 환자 진료중에 발생하는 임상자료와 분석도구를 의료진에게 제공하여 연구를 지원하는 시스템을 임상데이터웨어하우스(clinical data warehouse, CDW)라고 한다^{2,3}.

4. 당뇨병환자에서 저밀도콜레스테롤과 심혈관질환

당뇨병은 건강문제에 있어서 가장 중요한 만성질환의 하나로 최근 수년간 30세이상 성인에서 14%정도의 유병율을 보이고 있다⁴. 우리나라에서 2007년에 시행했던 연구에 따르면, 당뇨병환자의 원인별 사망중 심혈관질환이 암보다도 높은 것으로 나타나 심혈관 질환 예방의 중요성은 매우 크다고 할수 있다⁵. 하지만 당뇨병환자에서 철저한 혈당의 조절이 심혈관질환에 의한 사망의 위험을 낮추는 데는 효과적인지는 않을 것으로 나타난다. 오히려 심장질환을 예방하기 위해서는 여러 요인중에서도 LDL 콜레스테롤을 조절하는 것이 가장 중요함을 강조하는 연구결과도 존재한다⁶. 이러한 LDL 콜레스테롤 조절의 중요성 때문에 2001년에 발표되고 2004년에 업데이트된 ATPⅢ 가이드라인에서는 당뇨병을 심혈관질환과 동일한 위험으로 간주하고 LDL 콜레스테롤을 100미만으로 낮출 것을 권고하였다⁷. 이러한 진료지침은 약간의 변화도 있었지만 2015년 발표된 대한당뇨학회 당뇨병진료지침까지도 받아들여지고 있다⁸. 하지만 2016-2018년에 조사된 국민건강영양조사에 따르면 30세이상 성인 당뇨병환자의 72%가 고지혈증을 동반하고 있었으나 100미만으로 조절율은 52%밖에 되지 않았다⁴.

5. 국민건강보험 일산병원 연구정보시스템과 연구의 필요성

국민건강보험 일산병원은 최근에 미래를 위한 정보화 사업에 힘입어 CDM과 CDW를 모두 구축하였으며, 공단자료와 연계하여 연구에 활용할 수 있다는 연구에 있어 중요한

이점을 가지고 있다. 하지만 아직까지 구축 후에 연구를 통한 적절한 검증이 이루어지지 않은 상황으로 비교적 수준에 있는 심혈관질환 발생에 대한 후향적 코호트연구를 CDM과 공단자료를 연계한 DW자료, 본원 CDW 자료를 통해서 진행해보고 이를 통해서 현 데이터 베이스들의 현황과 문제점들을 파악하는 기회로 삼고자 한다.

제2절 연구 목적

이에 본 연구에서는 일산병원 CDM과 공단자료를 연계한 DW 자료 활용하여 일산병원 당뇨병환자 코호트를 구성하고 이에서 심혈관질환이 없는 당뇨병환자에서 LDL콜레스테롤을 낮추는 statin 제제의 복용여부와 이에 따른 LDL 콜레스테롤의 저하정도를 알아보고자 한다. 또한 추적관찰을 통해서 LDL 수치 및 스타틴 제제의 복용여부에 따른 심혈관질환의 발생위험 및 사망위험을 확인해 보고자 한다.

또한, 이 연구과정과 각 변수들이 연구에서 다루어지는 과정을 통해서 활용되는 데이터베이스의 현황과 연구과정에서의 문제점들을 통해 향후 연구자들이 해당 자료를 활용할 때 고려해야할 사항들에 대해서도 확인하여 추후 CDM CDW를 활용한 연구자들의 연구진행에 도움이 되고자 한다.

- 당뇨병환자의 statin 사용여부에 따른 LDL-C 수치의 변화
- 당뇨병환자의 statin 사용여부에 따른 심혈관질환 및 사망 위험
- 당뇨병환자의 statin 사용여부 및 LDL-C그룹에 따른 심혈관질환 및 사망위험

제2장

이론적 고찰

제1절 보건의료 정보화 실태조사 현황	19
제2절 국민건강보험 일산병원 임상데이터 웨어하우스	21
제3절 국민건강보험 일산병원 공통데이터 모델	24
제4절 당뇨병환자에서 저밀도콜레스테롤 조절과 심혈관질환	27

제2장 이론적 고찰

제1절 보건의료 정보화 실태조사 현황

한국보건의료정보원에서 2015-2017년까지 전국 574개 상급종합병원 및 종합병원을 대상으로 정보화현황 및 수준에 대한 실태조사를 실시하였다. 이 조사를 통해서 병원에 구축된 정보화기반 및 현황, 진료 및 연구 활용체계 등에 대한 전반적인 조사가 이루어졌다⁹⁾.

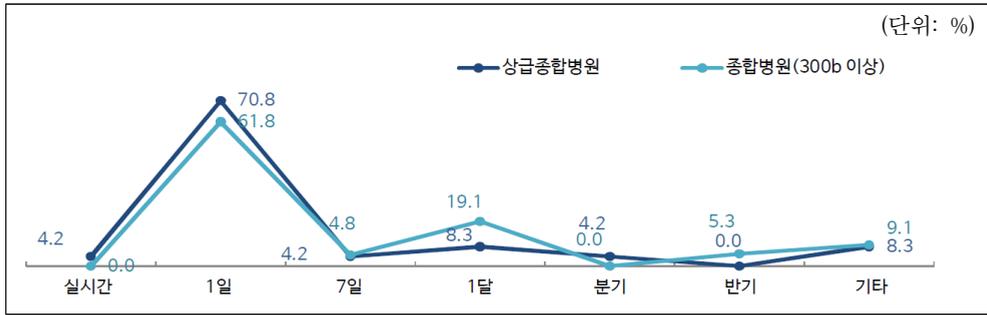
OECD 국가의 전자의무기록(EMR) 도입률은 일차 의료 81%, 입원 진료 76% 수준(2016년 기준)으로, EMR은 환자 안전, 의료 질 향상을 넘어 질병 관리 및 건강체계 향상을 위한 국가 보건의료정보화의 기반으로 인식되고 있음이 조사를 통해서 의료기관의 전자의무기록(EMR), 처방전달시스템(OCS), 영상정보시스템(PACS) 적용률은 높은 수준으로 조사되었다. 다수 의료기관은 모든 의무기록을 EMR로 사용(64.8~85.7%)하고 있으며, EMR과 수기기록을 병행하고 있는 경우까지 포함하면(96.5%~100%) 대부분 의료기관이 EMR을 사용하는 것으로 파악된다.

하지만 진료와 이를 지원하는 행정을 목적으로 구축된 EMR 시스템은 풍부한 데이터에 비해서 분석이나 데이터 추출을 위한 구조와 기능에 최적화되어 있지 않다. 이에 환자 데이터를 기반으로 한 연구 및 2차 분석을 위한 연구 정보시스템을 구축하고 있다. 특히 이번 조사에서 임상연구를 위한 윤리 지침 및 법적 규제를 시행하기 위해 익명화 기술 등을 임상데이터웨어하우스에 적용할 수 있으며, 다기관 연구를 위한 분산 연구망 지원을 위해 공통데이터모델(common data model, CDM)을 구축하고 있음을 확인하였다.

상급종합병원의 경우, 임상데이터웨어하우스(CDW), 공통데이터모델(CDM), 임상시험관리시스템은 50% 이상에서, 유전체분석정보시스템, 바이오은행관리시스템은 20~30% 수준에서 구축되어 있으나, 그 외 종합병원의 연구정보시스템 구축은 10% 내외 낮은 수준으로 구축되어 있다.

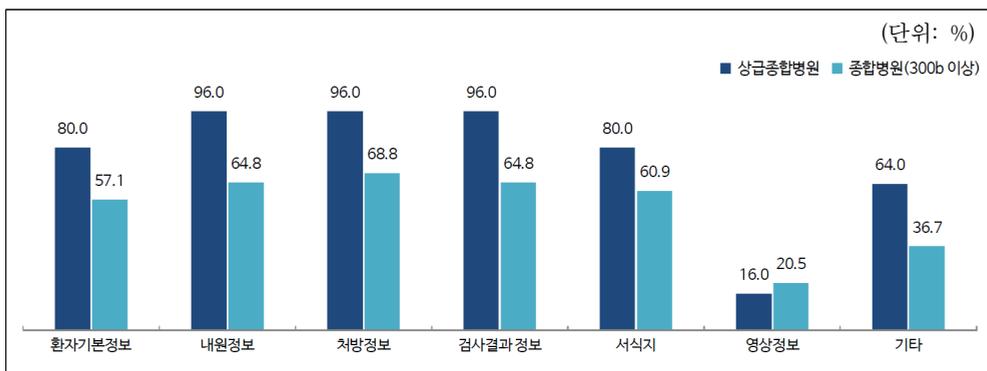
CDW와 CDM은 상급종합병원의 60% 수준에서 도입되어 있으며, 구축방식은 상용

솔루션보다 자체구축 비율이 높았다. CDW의 경우에 상급종합병원은 66.7%, 종합병원은 49.1%에서 CDW의 연구자 편의 기능인 웹인터페이스를 구축하고 있었으며, 데이터의 활용에 중요한 CDW의 ETL 주기는 1일이라고 응답한 의료기관이 상급종합병원의 70.8%, 300병상 이상 종합병원의 61.8%으로 가장 큰 비중을 차지하였다.[그림 2-1]



[그림 2-1] CDW의 ETL(데이터추출) 주기²⁾

사용자의 활용에 있어 중요한 CDW WEB interface 유무에 대해서는 상급종합병원 66.7%, 300병상 이상 종합병원 49.1%에서 제공을 하고 있었다. CDW의 제공데이터 종류는 상급종합병원의 경우 내원정보, 처방정보, 검사결과정보에서 모두 90% 이상의 병원에서 제공하고 있었으며, 환자기본정보, 서식지에 대해서 80.0% 제공하고 있었다. 반면에 종합병원은 내원정보, 처방정보, 검사결과정보, 서식지에서 60% 이상의 수준의 병원에서 제공하고 있었다.



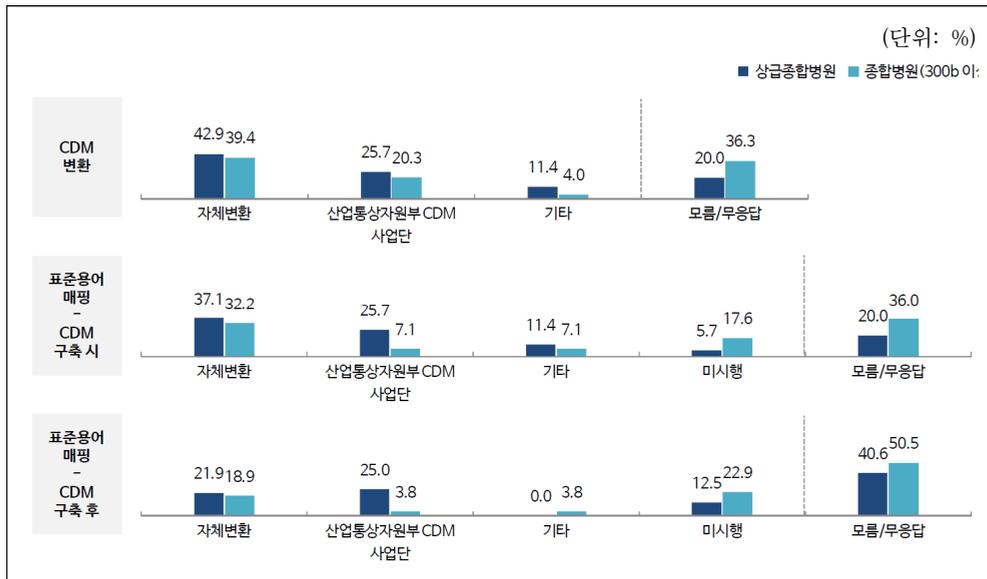
[그림 2-2] CDW의 제공데이터 종류³⁾

2) 출처: 2020년 보건의료정보화 실태조사 결과보고서

3) 출처: 2020년 보건의료정보화 실태조사 결과보고서

CDM의 경우에 대부분 OMOP CDM으로 상급종합병원의 69.4%, 종합병원의 32.1%에서 구축되어 있고, 약물 부작용 감시를 위한 K-CDM 도입은 이보다 훨씬 낮은 수준인 상급종합 11.8%, 종합병원 2.3%에서만 도입되어 있었다.

의료기관 데이터의 CDM 포맷으로의 변환은 산업통상자원부 CDM 사업단을 통하는 경우보다 자체변환이 많았고, 표준용어 매핑은 CDM 구축 시에는 자체변환이, CDM 구축 후에는 산업통상 자원부 CDM 사업단이 많았다.



[그림 2-3] CDM 변환 및 표준용어 매핑³⁾

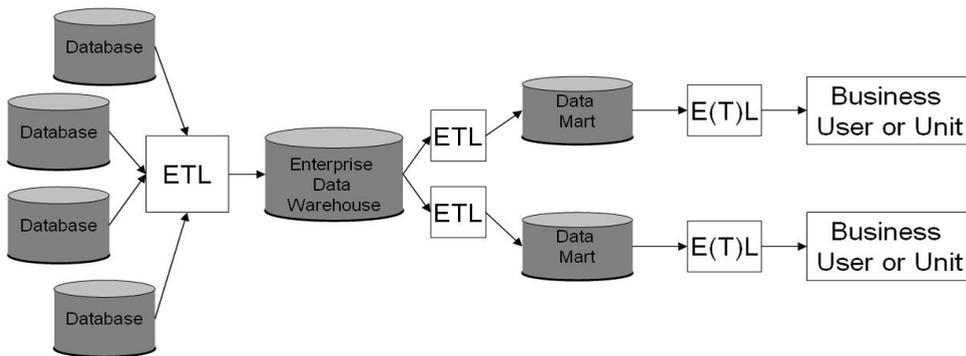
제2절 국민건강보험 일산병원 임상데이터웨어하우스

1. 데이터 웨어하우스 (Data Warehouse, DW)

DW는 사용자의 의사 결정에 도움을 주기 위하여 기간시스템의 데이터베이스에 축적된 데이터를 진료, 경영, 관리 등의 주제별 재구성 데이터를 관리하는 데이터베이스를 말한다. 정보(data)와 창고(warehouse)의 의미가 합성되어 만든 어휘로 방대한 조직 내에서 분산 운영되는 각각의 데이터 베이스 관리 시스템들을 효율적으로 통합하여 조정·관리함으로써 효율적인 의사 결정 시스템을 위한 기초를 제공한다.

3) 출처: 2020년 보건의료정보화 실태조사 결과보고서

데이터 웨어하우스의 구성은 관리 하드웨어, 관리 소프트웨어, 추출·변환·정렬 도구, 데이터 베이스 마케팅 시스템, 메타 데이터(meta data), 최종 사용자 접근 및 활용 도구 등으로 구성되어 있다. 특히 병원에서는 컴퓨터 성능의 발전으로 대규모 데이터 관리가 가능해지고 의료기관 데이터가 통합의료데이터베이스로 축적되었으며 의학연구, 병원 지식경영의 필요성이 증대되면서 DW의 필요성이 더욱 대두되었다. 그래서 DW는 주제 중심적, 통합적, 시계열적, 비휘발적(nonvolatile)인 자료의 집합으로 이야기된다¹⁰. [그림 2-4]



[그림 2-4] 데이터 웨어하우스의 개요도⁴⁾

2. 임상데이터웨어하우스(Clinical Datawarehouse, CDW)

2000년대 이후에 많은 병원들에서 DW가 구축되어 있긴 하나 대부분은 경영과 관련해서 경영정보, 원무, 검사 및 진료지원 등과 같이 행정적인 부분과 더욱 관련이 깊었다. 하지만 정밀의료, 의료빅데이터 등 보다 대규모의 의료정보를 통해서 환자에게 맞는 의료를 제공하고자 하는 요구가 많아지면서 진료중심의 데이터를 분석하여 환자와 의사에게 필요한 정보를 제공하는 CDW가 필요하게 되었다.

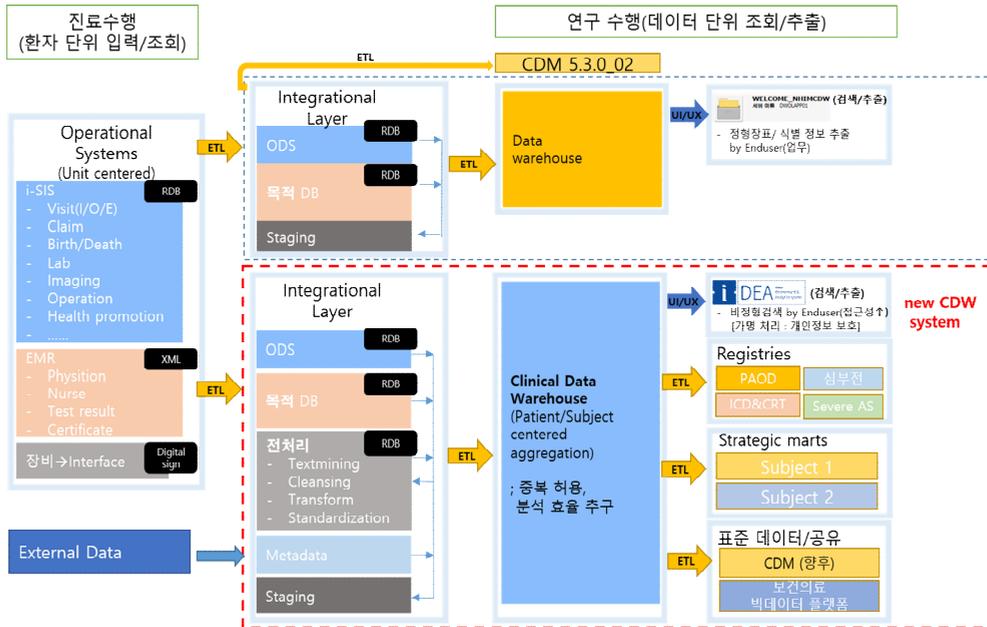
기존의 DW에서는 수치화된 데이터만을 주로 취급하였으나 최근에 병원에서 활용되기 시작하고 있는 CDW에서는 기존의 DW상의 데이터 이외에도 상세데이터와 텍스트 등 비정형데이터의 추출을 활용한 데이터들의 비교분석이 가능하다는 것이 특징이다. 또한, 특정한 SQL 프로그램을 통해서만 데이터를 추급이 가능했으나 CDW에서는 웹기반의 인터페이스를 설치융합함으로써 연구자들이 보다 쉽게 데이터를 활용할 수 있도록 하는 것도 최근 CDW의 특징 중 하나라 하겠다.

4) 출처: <https://ko.wikipedia.org/wiki/>

3. 국민건강보험 일산병원 DW 및 CDW

본 국민건강보험 일산병원은 21세기 의료환경은 의료욕구 증가, 노인인구 증가, 질병 구조의 변화에 따라 새로운 의료서비스 환경과 병원운영모델 제시가 절실히 요구되고 있고, 공익적 차원에서 새로운 의료서비스와 운영체계를 개발·적용할 보험자 직영병원의 필요와 의료서비스 취약지역인 경기서북부 지역에 양질의 의료서비스를 제공하기 위하여 경기도 고양시 일산에 병상수 745병상 규모로 (현 822병상) 2000년 3월3일 개원하면서 초기부터 DW를 구축하여 이용하여 왔다. 하지만 2011년 전자의무기록시스템을 구축하면서 시스템 변경과 함께 새로운 DW를 구축하여 사용중이었다.

최근 좀더 고도화된 DW시스템의 요구가 대두되면서 2019년부터 새로운 CDW 시스템 구축을 시행하게 되었다.



[그림 2-5] 국민건강보험 일산병원 DW와 CDW의 모식도

기존의 구DW는 전자의무기록시스템으로부터 정보를 추출받아서 주제별로 정리된 것으로 정보의 사용 및 검색이 매우 제한적으로 이루어졌다. 해당 데이터베이스를 바탕으로 현재 사용중인 국민건강보험 일산병원 CDM 5.3.0 버전도 구축되었다.

하지만, 새롭게 구축된 CDW에서는 기존의 DW에서 제한되었던 기능에 비해서 추출

작업이 빠르게 진행되고 텍스트 마이닝과 같은 비정형데이터에 대한 처리가 가능하게 되었으며 외부자료와의 연결 및 레지스트리 구성과 같은 새로운 기능이 추가되었다. [그림2-5] 뿐만 아니라 연구자들이 보다 손쉽게 연구 및 진료에 사용가능하도록 검색조건 등을 활용하여 환자를 검색하고 레지스트리 구성 등의 기능을 활용할수 있는 웹기반의 인터페이스인 iDEA를 함께 운용하게 되었다. 향후 CDW ready CDM이라는 시스템으로 CDW 데이터베이스가 매일 추출된 자료로 업데이트될 때 자동적으로 CDM 데이터가 함께 업데이트되는 시스템까지 구현할 계획을 가지고 있다.

제3절 국민건강보험 일산병원 공통데이터모델

1. 공통데이터 모델(Common data model, CDM) 과 Observational Medical Outcomes Partnership (OMOP)-CDM

관찰연구에서 자료 수집을 위해 비용과 시간면에서 상당한 제약이 있는 전향적인 연구 대신에 기존에 존재하는 자료를 바탕으로 시행하는 후향적인 연구가 각광을 받고 있다. 하지만, 이러한 자료를 바탕으로 한 후향적 연구는 단일 기관의 데이터만으로는 일반화할 수 있는 결과를 도출하기에는 어려움이 존재한다. 또한 국민건강보험공단 자료나 건강심사평가원 자료와 같은 큰 규모의 자료도 존재하지만 이러한 데이터베이스는 보험청구자료를 기반으로 하고 있어 보다 자세한 임상자료를 확보하는데는 어려움이 있다.

따라서, 보다 풍부한 임상자료를 바탕으로 하되 서로 다른 데이터 출처로부터 일반화된 연구결과를 도출하기 데이터를 동일한 데이터 표준으로 변환할 필요성을 대두되었으며, 이를 공통데이터모델을 구현함으로써 가능해지게 되었다.

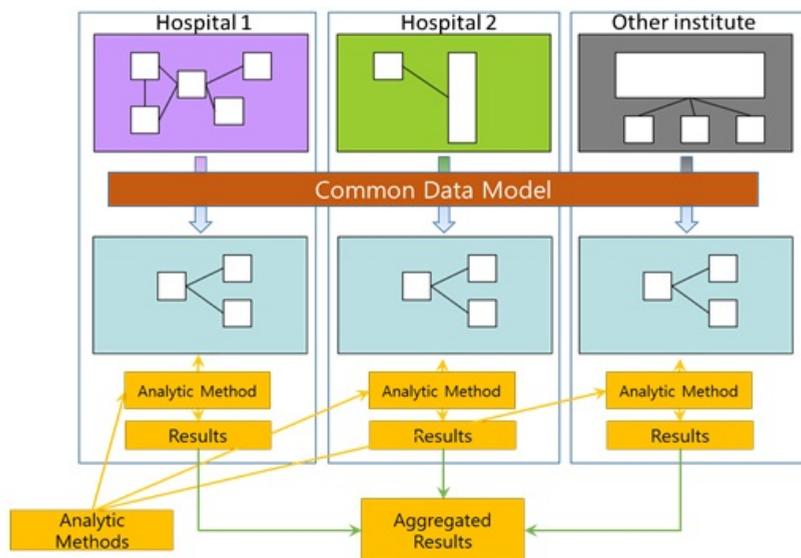
또한, 개인정보 등과 같은 민감한 문제들을 해결하는 방법으로 분산연구망을 구축하는 방안이 모색되었고, 이러한 예들이 세계적으로 다양하게 나타나게 되었다. Exploring and Understanding Adverse Drug Reactions (EU-ADR), European Medical Informatics Framework, Vaccine Safety Datalink, Food and Drug Administration (FDA)의 Sentinel, Patient-Centered Outcomes Research Network (PCORnet) 등 다양하게 존재한다¹¹⁻¹³.

그 중에서도 Observational Medical Outcomes Partnership (OMOP)-CDM은 표준화된 데이터 구조와 분석 가능한 변수들의 집합을 정의하기 위해서 다양한 규칙을 적용하

24 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

는 데이터 매핑 모델을 사용하고 있다¹⁴. OMOP-CDM은 미국 식품의약국 FDA이 주관하고, 미국 국립 보건원(National Institutes of Health) 관리하에 학술연구자, 보건데이터 파트너 및 협력 제약사간의 컨소시엄으로 구성되었으며, 보건의료 데이터를 이용하여 약물에 대한 능동적 감시 체계를 만들고자 시작하였다. 이후 다수의 청구자료 및 전자의 무기록 데이터 베이스에 적용하여 대안적인 역학 설계 및 통계 방법의 성능을 검증하는 실험을 설계하였다. 이후 연구진들은 데이터의 구조, 내용 및 용어를 표준화하여 하나의 통계분석코드가 모든 데이터 파트너에서 공통으로 사용되도록 OMOP-CDM을 설계하였다¹⁵.

OMOP CDM은 다양하게 주어진 원천데이터로부터 데이터세트를 구성하고, 전체 범위의 의료이용 구조를 포함하며 다양한 변수들을 수록한다. 단, 매핑시 원자료의 용어나 정보대신 표준화되어 매핑된 데이터를 사용하도록 설계되어 있다. 이러한 특징 때문에 한번 의료정보가 한번 CDM으로 변환되면 고정적인 데이터가 되어 연구자가 지속적으로 표준화된 데이터의 활용이 가능하게 된다. 이렇게 각 기관에서 구성된 데이터베이스를 바탕으로 공동데이터모델을 위한 네트워크 안에서 분산형 연구망이 만들어지고 이를 통하여 데이터자체가 아닌 데이터 분석프로세스만을 오가는 방법으로 각 기관에서 분석이 가능하도록 하고 있다.



[그림 2-6] 분산형 연구망⁵⁾

5) 출처: <http://www.feedernet.org/> 분산형바이오헬스 빅데이터 사업단

2. 분산형 바이오헬스 빅데이터 사업

OMOP-CDM의 연구 성과는 Observational Health Data Sciences and Informatics (OHDSI) 개방형 과학 공동체로 이어졌으며¹⁶, 한국에서는 2014년 아주대학교병원을 시작으로 전자의무기록의 OMOP-CDM으로 변환이 이루어졌다. 이후 OHDSI 및 CDM의 의미가 전파되면서 2018년 산업통상자원부가 지원하는 분산형 바이오헬스 빅데이터 (Federated E-Health Big Data for Evidence REnovation Network) 사업으로 한국 내에서 OMOP-CDM를 바탕으로 하는 분산형 연구망을 구축하는 사업이 진행하게 되었다. 2022년 현재 전국적으로 62개의 상급종합병원 및 종합병원을 데이터 파트너로 사업을 진행하고 있다.

3. 국민건강보험 일산병원 CDM

국민건강보험 일산병원도 분산형 바이오헬스 빅데이터 사업의 일환으로 2018년 9월부터 병원에 보유한 임상데이터를 OMOP_CDM 5.3 버전 기반의 데이터 베이스로 변환하는 사업을 진행하였다.

이렇게 일산병원 CDM인 NHIMC 5.3.0_01 버전이 1,357,014명을 대상으로 구축되었다. 2022년 현재는 3개의 버전이 사용가능한 상태이다. 5.3.0_02버전은 초기 01버전을 고도화한 결과물로 매핑코드가 변환사업을 시행한 업체로부터 공개되어 있는 상태이다. 최근에 일산병원 CDW가 새롭게 구축되면서 이를 바탕으로한 새로운 CDM 5.3.1_03버전이 구축되었다. 기존의 01,02 버전은 일산병원의 구 DW로부터 원천자료를 받았던 것과는 달리 03버전은 장기적으로 CDW로부터 원천자료를 받을수 있도록 설계되었으며 현재는 일부자료는 구 DW로부터 일부는 새로운 CDW로부터 받도록 설계되었으며, 향후 CDW를 고도화하면서 모든 자료를 CDW에서 받을수 있도록 할 예정이다. 5.3.1 weekly 버전은 03버전을 매주 토요일 적재를 통해서 실시간에 가깝게 업데이트하는 버전이다<표 2-1>.

26 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

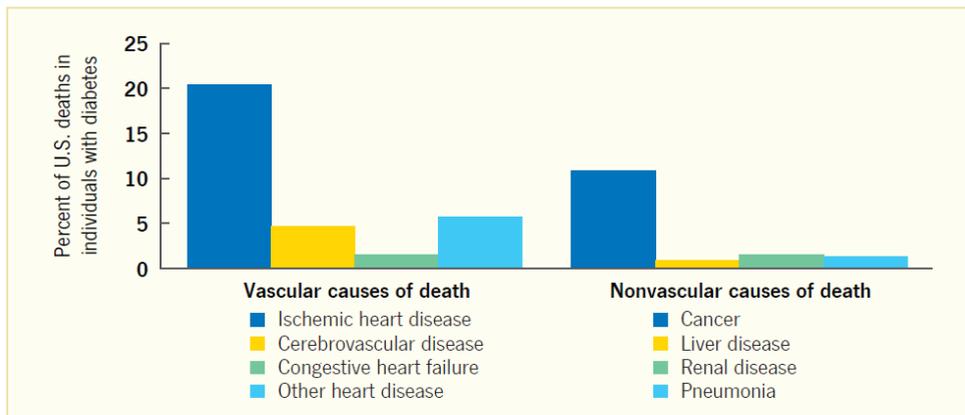
<표 2-1> 국민건강보험 일산병원 버전별 CDM 데이터베이스 특징

기준	DATA version		
	5.3.0_02	5.3.1_03	5.3.1_weekly
환자수	1.367M	1.437M	1.437 M
Follow up 기간	2000~2020	2000~2022	2000~2022
현재 사용가능여부	O	O	O
ETL 상황	2회/년	-	매주 토요일 적재
Mapping 코드 공개여부	O	X	X
DATA source	본원 DW	본원 DW, CDW	본원 DW, CDW
특성	5.3.0_01 초기버전의 고도화 결과	CDW 연결된 개발버전	

제4절 당뇨병환자에서 저밀도콜레스테롤 조절과 심혈관질환

1. 당뇨병환자에서 심혈관질환 위험

2008년 미국에서 2형 당뇨병환자를 대상으로 확인된 사망원인을 보면 모든 원인중에서 허혈성 심장질환이 가장 높은 비중을 차지하고 있다¹⁷.

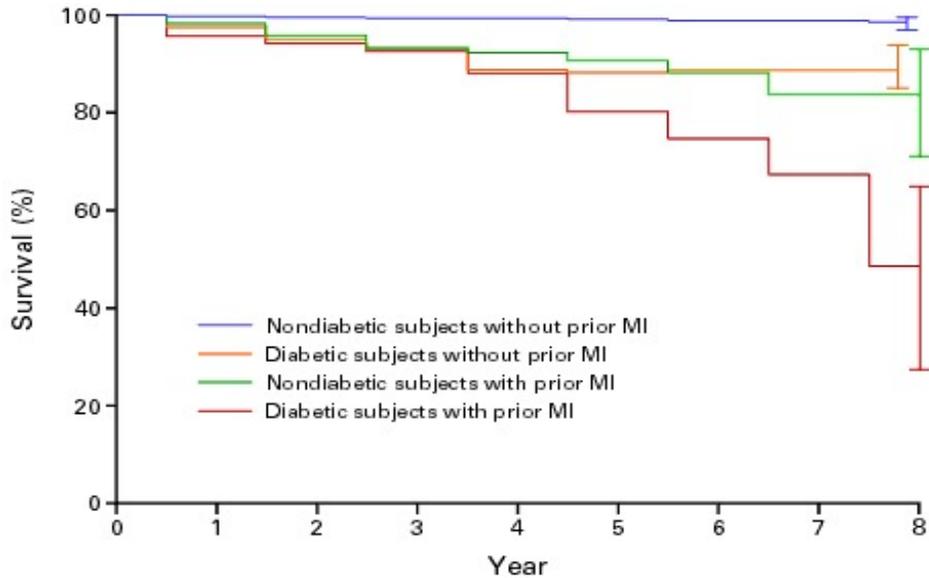


SOURCE: National Vital Statistics System 2008

[그림 2-7] 2008년 미국 당뇨병환자에서의 사망원인⁶⁾

6) 출처: In: Diabetes in America. 3rd edition. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (US); 2018 Aug. CHAPTER 36.

또한, 핀란드 인구에서 당뇨병을 가진 대상자와 당뇨병이 없는 대상자군에서 심근경색의 발생을 본 연구에서 이전에 심근경색이 있었던 당뇨병이 없는 환자와 심근경색의 과거력이 없는 당뇨병환자의 심근경색 발생위험이 비슷하게 나타났다¹⁸. 이 연구를 바탕으로 2004년 업데이트된 ATP(Adult Treatment Panel) III 가이드라인에서는 당뇨병을 높은 위험의 허혈성심질환 위험요인으로 간주하였다⁷.



[그림 2-8] 당뇨병환자 및 비당뇨병환자에서 심근경색 사망에 대한 위험에 대한 생존분석 곡선⁷⁾

2. 당뇨병환자에서 저밀도콜레스테롤 조절과 심혈관질환위험

실제 여러 연구로부터 당뇨병환자에서 심혈관질환의 위험을 감소시키기 위해서 저밀도콜레스테롤(LDL-Cholesterol, LDL-C)를 조절하는 것이 심혈관질환의 위험을 낮출수 있는 것으로 나타났다.

1998년 BMJ에 발표된 UKPDS 연구에서 당뇨병환자에서 여러 관상동맥질환의 위험인자들에 대해서 조사한 결과 LDL-C이 가장 중요한 위험인자로 확인되었으며 LDL-C를 1mmol/L을 줄일 때 관상동맥질환의 36%를 감소시킬수 있는 것으로 나타났다⁶.

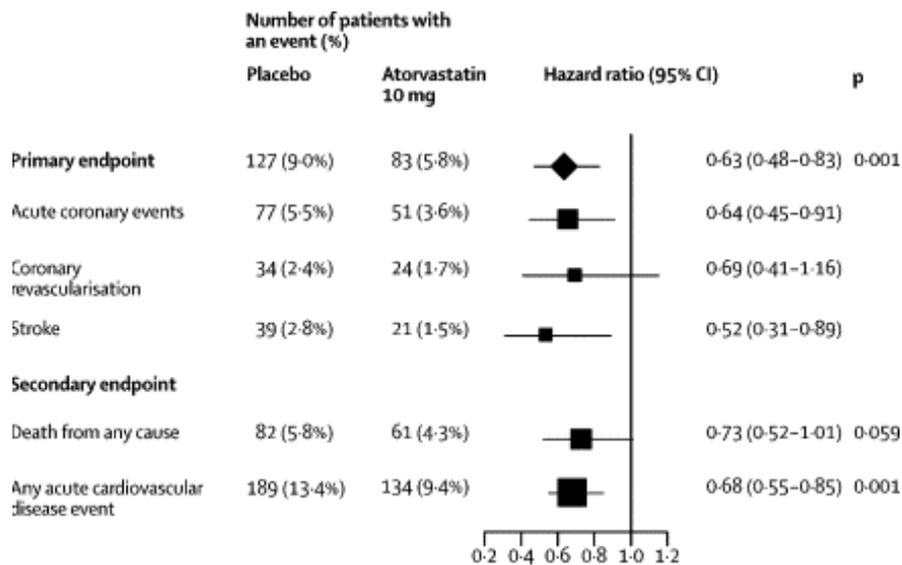
특히 2004년에 Lancet지에 발표되었던 CARDS 연구에서는 영국에서 2,838명의 당뇨병환자를 대상으로 위약군과 statin제제인 atorvastatin 10mg를 복용한 군에서 심혈관질환

7) 출처: DOI: 10.1056/NEJM199807233390404

28 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

환의 발생위험을 확인하였다¹⁹. 3.9년의 평균 추적기간후에 statin 복용군에서 위약군에 비해서 적어도 한가지 이상의 심혈관질환의 위험이 37% 감소하였다. 또한 급성심근경색은 36%, 관상동맥재관류는 31% 전체사망위험도 27% 감소하였다.

2013년에 발표된 연구에서는 26,000명의 당뇨병환자를 대상으로 당화혈색소, 수축기혈압, LDL-C의 조절정도와 심혈관질환으로 인한 입원을 살펴본 연구에서 수축기혈압을 130mmHg 미만으로, LDL-C 100mg/dl 미만으로 조절한 군에서 각 요인들을 단독으로 또는 당화혈색소의 조절이 잘된 다른 조합의 군에서 비해서 심혈관질환에 의한 입원의 발생이 적었다²⁰.



[그림 2-9] Atrovastatin 사용여부에 따른 심혈관질환 및 사망위험⁸⁾

2019년에 발표된 기저의 심혈관질환이 없는 한국인 당뇨병환자를 대상으로한 연구에서도 statin사용자군, 비사용자군 모두에서 LDL-C 수치 높은 군에서 심근경색의 위험이 상승하였으며, 특히 statin 비사용군에서는 LDL-C 100mg/dl 이상군에서 상승하였다²¹.

8) 출처: DOI: 10.1016/S0140-6736(04)16895-5

3. 당뇨병환자에서 이상지질혈증 조절 지침

여러 연구결과를 바탕으로 2001년에 제정되어 2004년에 업데이트되었던 ATP III 지침에서는 당뇨병을 다른 허혈성심질환위험과 동일한 위험으로 간주하고 100mg/dl 미만으로 조절할 것을 권고하고 있다.

이러한 지침은 이후 미국 유럽 지침들에 의해 조금씩 변화가 있었으나 대한 당뇨병학회에서는 2013, 2015년 당뇨병 진료지침에서도 심혈관질환이 없는 당뇨병환자에서는 LDL-C를 100mg/dl 미만으로 조절할 것으로 지속적으로 권고하고 있다⁸.

<표 2-2> ATP III 가이드 라인⁹⁾

TABLE 2. ATP III LDL-C Goals and Cutpoints for TLC and Drug Therapy in Different Risk Categories and Proposed Modifications Based on Recent Clinical Trial Evidence

Risk Category	LDL-C Goal	Initiate TLC	Consider Drug Therapy**
High risk: CHD* or CHD risk equivalents† (10-year risk >20%)	<100 mg/dL (optional goal: <70 mg/dL)‡	≥100 mg/dL#	≥100 mg/dL†† (<100 mg/dL: consider drug options)**
Moderately high risk: 2+ risk factors‡ (10-year risk 10% to 20%)§§	<130 mg/dL¶	≥130 mg/dL#	≥130 mg/dL (100–129 mg/dL; consider drug options)‡‡
Moderate risk: 2+ risk factors‡ (10-year risk <10%)§§	<130 mg/dL	≥130 mg/dL	≥160 mg/dL
Lower risk: 0–1 risk factors§	<160 mg/dL	≥160 mg/dL	≥190 mg/dL (160–189 mg/dL: LDL-lowering drug optional)

*CHD includes history of myocardial infarction, unstable angina, stable angina, coronary artery procedures (angioplasty or bypass surgery), or evidence of clinically significant myocardial ischemia.

†CHD risk equivalents include clinical manifestations of noncoronary forms of atherosclerotic disease (peripheral arterial disease, abdominal aortic aneurysm, and carotid artery disease [transient ischemic attacks or stroke of carotid origin or >50% obstruction of a carotid artery]), diabetes, and 2+ risk factors with 10-year risk for hard CHD >20%.

‡Risk factors include cigarette smoking, hypertension (BP ≥140/90 mm Hg or on antihypertensive medication), low HDL cholesterol (<40 mg/dL), family history of premature CHD (CHD in male first-degree relative <55 years of age; CHD in female first-degree relative <65 years of age), and age (men ≥45 years; women ≥55 years).

§§Electronic 10-year risk calculators are available at www.nhlbi.nih.gov/guidelines/cholesterol.

§Almost all people with zero or 1 risk factor have a 10-year risk <10%, and 10-year risk assessment in people with zero or 1 risk factor is thus not necessary.

¶Very high risk favors the optional LDL-C goal of <70 mg/dL, and in patients with high triglycerides, non-HDL-C <100 mg/dL.

‡Optional LDL-C goal <100 mg/dL.

#Any person at high risk or moderately high risk who has lifestyle-related risk factors (eg, obesity, physical inactivity, elevated triglyceride, low HDL-C, or metabolic syndrome) is a candidate for therapeutic lifestyle changes to modify these risk factors regardless of LDL-C level.

**When LDL-lowering drug therapy is employed, it is advised that intensity of therapy be sufficient to achieve at least a 30% to 40% reduction in LDL-C levels.

††If baseline LDL-C is <100 mg/dL, institution of an LDL-lowering drug is a therapeutic option on the basis of available clinical trial results. If a high-risk person has high triglycerides or low HDL-C, combining a fibrate or nicotinic acid with an LDL-lowering drug can be considered.

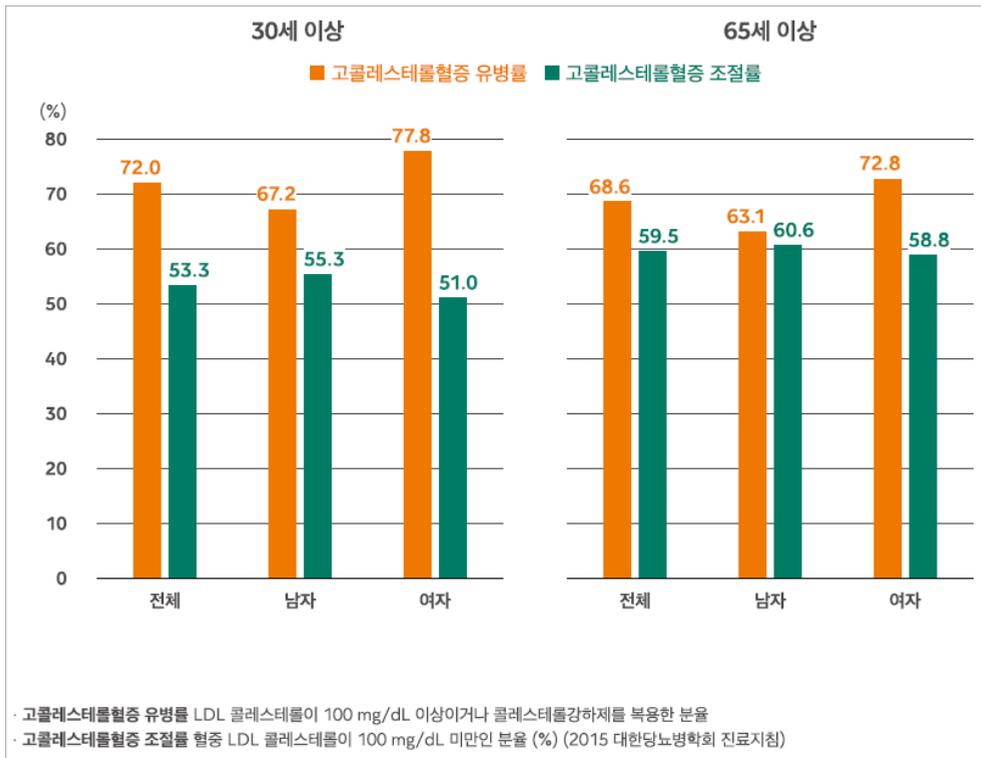
‡‡For moderately high-risk persons, when LDL-C level is 100 to 129 mg/dL, at baseline or on lifestyle therapy, initiation of an LDL-lowering drug to achieve an LDL-C level <100 mg/dL is a therapeutic option on the basis of available clinical trial results.

4. 당뇨병환자에서 이상지질혈증의 조절을

2020년에 대한당뇨병학회에서 발행한 Diabetes Fact Sheet 2020 에 따르면 국민건강영양조사결과를 바탕으로 분석하였을 때, 30세이상에서 13.8%에서 당뇨병을 가지고 있었다. 그중에서 LDL-C를 100mg/dl 이상이거나 콜레스테롤강하제를 복용하는 비율이 72%에 해당되었으며, 반면에 당뇨병 유병자 중 55.3% 만이 LDL-C 수치가 목표치대로 조절되었다⁴.

9) 출처: DOI: 10.1161/01.CIR.0000133317.49796.0E

30 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교



[그림 2-10] 당뇨병환자에서 고콜레스테롤혈증 유병률과 조절률¹⁰⁾

5. Statin제제와 임상적 실제사용의 문제

많은 연구결과들과 진료지침들을 통해서 당뇨병환자에서 LDL-C의 조절은 환자를 보는 임상 의들에게도 당뇨병 환자들의 심혈관 질환위험을 줄이기 위해 매우 중요한 문제로 인식되고 있다. 따라서 환자를 관리함에 있어 당뇨진단시 및 정기적으로 LDL-C의 수치를 측정하고 LDL-C 수치가 높은 경우에는 어려움없이 statin제제를 처방하여 이를 조절하는 경우가 많다. 하지만 LDL-C의 수치가 100mg/dl를 약간 상회하는 정도로 조절되는 경우에는 statin제제의 사용을 주저하는 경우도 발생하는 것이 현실적인 처방의 문제이기도 하다.

10) 출처: Diabetes Fact sheet in korea 2020

제3장

연구 내용 및 방법

제1절 연구 내용	35
제2절 연구 방법	39

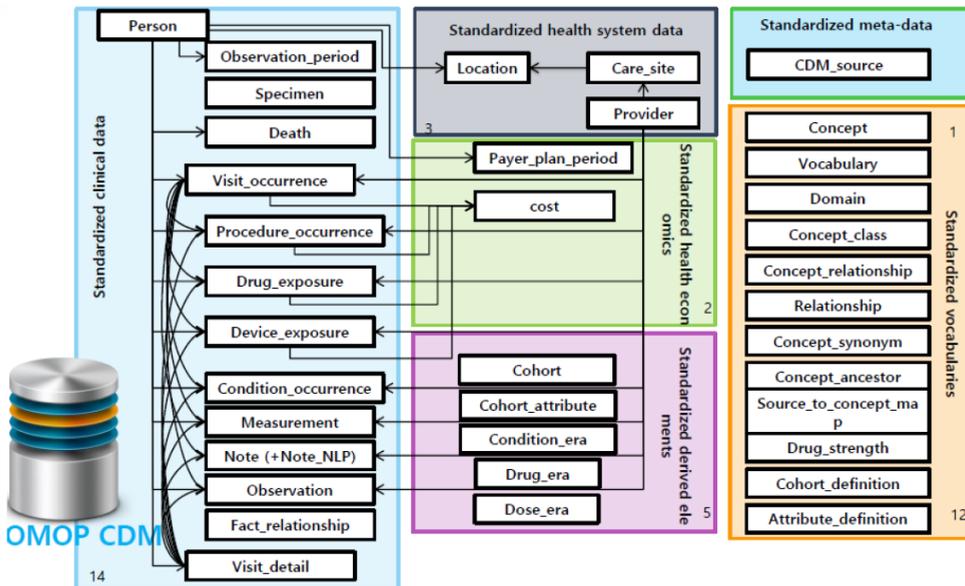
제3장 연구 내용 및 방법

제1절 연구 내용

1. 국민건강보험 일산병원 CDM 현황

1) 데이터질평가

현재 사용가능한 일산병원 CDM database에 대해서 연구시행의 관점에서 현황을 파악하고자 한다. 이전 DW로부터 추출된 데이터는 CDM의 도메인별로 해당되는 표준용어에 매핑되도록 구성되어 있다.[그림 3-1] 각 도메인(테이블)별로 표와 같이 해당 데이터의 내용을 확인할수 있다.[표 3-1] 이에 탑재된 레코드에 대한 매핑정도를 파악하였다.



[그림 3-1] OMOP-CDM 구조 개요도¹¹⁾

11) 출처: <http://www.feedernet.org/> 분산형 바이오헬스 빅데이터 사업단

<표 3-1> CDM 테이블(도메인)별 설명

CDM Table명	테이블 설명
Person	환자의 인종, 생년월일, 성별 등의 기본정보를 알 수 있는 테이블
Visit_occurrence	환자가 병원에 방문한 기록으로 방문시작일과 방문종료일을 알 수 있는 테이블
Observation_period	환자가 병원에 내원하여 치료가 완료되기까지의 기간을 알 수 있는 테이블
Procedure_occurrence	진단 및 치료 목적을 가지고 의사에 의해 처방 및 지시된 활동 또는 프로세스를 알 수 있는 테이블
Drug_exposure	환자에게 처방된 약 정보로 처방약, 약에 노출된 일시, 중지사유, 처방일수, 처방용량, 투약방법, 복용법, 단위 등을 알 수 있는 테이블
Condition_occurrence	환자의 진단명, 건강상태, 진단일시, 진단의사등을 알 수 있는 테이블
Device_exposure	기기 및 각종 치료재료 사용 정보를 알 수 있는 테이블
Measurement	체계적인 검사로부터 얻어진 측정값으로 측정일, 측정값 최대/최소 범위, 측정 값, 측정단위 등을 알 수 있는 테이블
Specimen	검체채취일시, 양, 위치, 단위등을 알 수 있는 테이블
Observation	문진, 과거력, 알리지 등의 정보를 알 수 있는 테이블
Death	환자의 사망일, 원인을 알 수 있는 테이블
Note	자유기술문으로 구성된 각종 기록지의 정보를 알 수 있는 테이블
Health System data	
Location	물리적 위치 또는 주소정보를 캡처하는 일반적인 방법을 나타냄
Care_site	의료서비스가 시행되는 조직 단위의 고유한 목록 포함 (사무실, 병동, 병원, 진료소 등)
Provider	의료 서비스제공자 식별 목록 포함
Meta-data	
CDM_source	원본 데이터베이스를 OMOP CDM으로 변환하는데 사용되는 소스에 대한 세부 정보 포함.(sourcename, holder, source description)

또한 데이터의 질을 확인할 수 있는 자체도구인 Achilles 및 자체분석도구인 ATLAS 내의 Databoard로부터 현재 데이터의 품질상태를 확인하였다.

2) 연구관련 변수별 매핑현황확인

현재 병원내에서는 병원자체코드를 사용하고 있으며 청구를 위해서는 건강보험코 드인 EDI 코드, ICD 코드와 유사한 한국표준질병사인분류코드인(KCD코드) 등의 다른 코드를 병행하여 사용하고 있다. 이에 주요변수들에 대해서 현재 CDM database에 대해 매핑상태를 확인하고자 하였다.

3) CDM database 혼란 변수 레코드 확인

CDM 데이터베이스 내에서 연구를 위해 사용가능한 혼란변수의 현황을 확인하 였다.

2. 국민건강보험 일산병원 CDW 현황

1) 구DW와 CDW 통한 산출 대상자수 비교확인

최근 새롭게 구축된 CDW 시스템은 구 DW로부터 데이터를 이관한 후 최근에 새롭게 발생하는 데이터를 추가적으로 확보하여 구축된 시스템이다. 이에 구DW

데이터베이스와 새로운 CDW의 데이터베이스를 비교하는 것은 추후 연구자들이 새로운 CDW 시스템을 통해서 연구를 진행함에 있어 매우 중요한 사항이다. 이에 구 DW와 새로운 CDW간에 동일한 연구대상자 조건에 따른 대상자수를 확인하여 이를 통해 현재 CDW의 데이터질의 현황을 확인하였다.

2) CDW 웹인터페이스 iDEA의 현황

연구자들이 연구를 진행함에 있어 기존에는 의무기록부, 정보팀 등의 병원내 정보 부서를 통해서 DW로부터 연구에 필요한 대상자를 산출하는 방식을 활용해왔다. 실제로 연구자들이 직접 DW나 CDW로부터 MSSQL, Sybase 등의 프로그램을 활용하여 직접 여러 가지 조건들을 통해 대상자를 산출하기는 쉽지가 않기 때문이다. 따라서 많은 병원에서 새로운 연구정보시스템으로 CDW를 구축하면서 연구자들이 쉽게 사용할 수 있는 웹기반의 인터페이스를 함께 구축하는 추세이다.

이에 일산병원 CDW의 구축과 함께 사용가능해진 웹인터페이스인 iDEA는 연구자들에 있어 가장 새로운 CDW시스템을 활용하는 가장 중요한 도구라 할 수 있다. iDEA는 CDW 데이터베이스로부터 대상자를 산출하는 기능을 주된 기능으로 하고 있어 특정 조건에 대하여 CDW로부터 직접 산출한 대상자수와 iDEA를 통해 산출한 대상자수를 비교함으로써 iDEA의 현재 기능 상태를 확인할 수 있다. 또한 실제로 연구자들이 대상자를 산출하는 조건을 당뇨병환자에서 LDL-C와 심혈관질환 연구에서의 예시를 통해서 현재의 iDEA의 기능의 제한점 및 한계를 확인할 수 있다.

3. CDM database 및 공단자료와 연계한 DW자료로부터 당뇨병환자에서 LDL-C과 심혈관질환 및 사망위험

1) CDM 데이터베이스를 기반으로 한 Statin 복용여부에 따른 심혈관질환의 발생위험 및 사망위험

일산병원 CDM 데이터베이스를 기반으로 Statin 복용여부에 따른 심근경색, 급성뇌졸중, 관상동맥재관류의 발생위험 및 사망위험에 대해서 알아보고자 한다.

2) DW 데이터베이스를 기반으로한 공단연계 자료로부터 당뇨병환자에서 Statin제제 복용여부에 따른 심혈관질환의 발생위험 및 사망위험

첫째로, Statin 사용군 및 비사용군에서 18개월이후에 LDL-C 및 non-HDL-C의 변화를 확인하였다. Statin 사용군에서 약물사용후에 LDL-C 수치가 안정화되는 18개

월 이후의 시점에서 LDL-C 수치를 Statin 비사용군에서 당뇨병진단후부터 18개월 이후의 시점과 비교하고자 한다. 이를 통해서 실제 statin 사용군과 비사용군에서의 LDL-C 및 non-HDL-C 감소율과 조절율을 알아보하고자 한다.

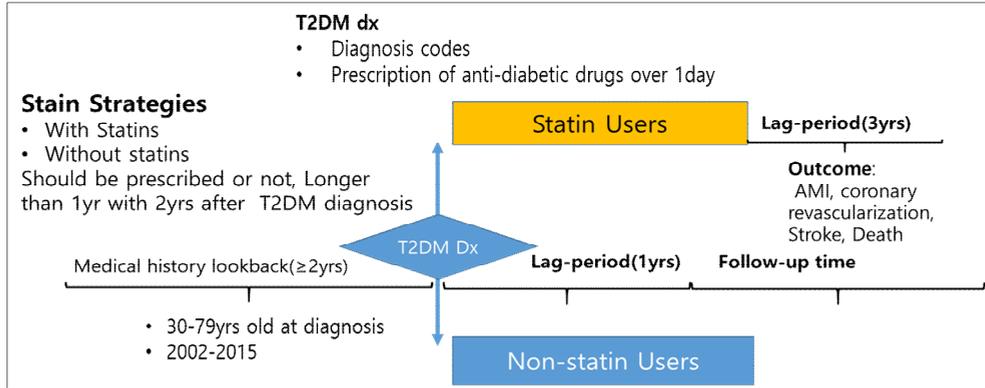
둘째, Statin 사용군과 비사용군에서 심혈관질환 발생 및 사망위험 비교

Statin제제 사용군과 비사용군에서 급성심근경색, 관상동맥 재관류, 급성뇌경색, 급성뇌출혈의 발생위험을 알아보하고자 한다. 또한 두 군에서 총 사망위험의 차이를 알아보하고자 한다.

셋째, Statin 사용군과 비사용군에서 LDL-C 수치에 따른 심혈관질환 발생 및 사망 위험 비교하고자 한다. Statin제제 비사용군의 LDL-C 100mg/dl 미만인 그룹을 참고군으로 하여 Statin제제 사용군의 LDL-C 수치에 따른 그룹별 심혈관질환 발생 및 사망위험을 비교해보고자 한다.

제2절 연구 방법

1. 연구 대상자



[그림 3-2] 연구의 틀

본 연구의 대상은 국민건강보험 일산병원에서 2002년 1월 1일부터 2025년 12월 31일까지 당뇨병으로 진단받은 환자 30세이상 79세이하의 환자를 대상으로 하였다. [그림 3-2]

2002년 1월 1일부터 2015년 12월 31일까지 2형 당뇨병환자를 진단받은 30-79세 성인 으로 하였다. 당뇨병의 진단은 당뇨병진단코드를 가지는 환자가 당뇨병약물을 1일이상 처방받은 것으로 정의하였다. 당뇨병진단코드는 한국표준질병사인분류(KCD)에서 2형 당뇨병에 속하는 E11,12,13,14 코드로 하였다. 또한 당뇨병으로 진단받은 사람중에서 다른 원인에 의한 LDL-C저하를 배제하기 위해서 당뇨병진단전 LDL-C이 70mg/dL 이상인 환자로 국한하였으며, 중성지방의 수치도 500mg/dL 미만인 환자로 국한하였다.

대상자의 제외기준으로 기저 및 치료중 당뇨조절수치가 Hba1c 12초과인 대상자를 당뇨병의 조절정도가 적절하지 않다고 판단하여 제외하였다. 심혈관질환의 일차예방을 위한 목적으로 연구를 진행하기 위해서 과거 심근경색, 협심증을 진단받았거나 관상동맥 중재술, 관상동맥우회술과 같은 관상동맥재관류술을 받은 과거 허혈성 심장질환자는 제외하였다. 또한 급성 뇌경색, 일과성 뇌허혈의 기왕력이 있는 경우에도 제외하였다. 당뇨병의 진단과 함께 Statin제제를 통한 심혈관질환의 위험을 확인하기 위해서 기저에 statin제제, ezetimibe제제 및 fibrate 제제를 복용하는 환자는 제외하였다.

Statin제제 사용군과 statin제제 비사용군으로 구분하였으며, statin제제 사용군은 당뇨병진단후 2년 이내에 1년이상 statin제제를 처방받은 환자를 statin제제 사용군으로 정의

하였으며 Statin제제 비사용군은 당뇨병진단후에 statin제제를 사용하지 않은 군으로 정의하였다.

2. 변수의 정의

1) 대상자 관련 변수

<표 3-2> 대상자 관련 변수 및 관련내용

	용도	요인	KCD 코드	기타코드	추가정보
1	2형당뇨병 정의	당뇨병진단(진료)	E11-14		병원내 첫진단
		당뇨병약제		당뇨병약물코드	1일 이상 처방.
2	연령	진단시30-79세			진단시 나이
3	진단일자	2002.01.01-2015.12.31			
4	기저 콜레스테롤 수치	LDL-cholesterol > 70mg/dl		검사코드 및 수치	당뇨병진단전 6개월이내
5		Triglyceride < 500mg/dl		검사코드 및 수치	당뇨병진단전 6개월이내
6		Myocardial infarction	I21, I22		당뇨병진단전
7		Placement of coronary stent (PCI)		관상동맥중재술코드	당뇨병진단전
8	과거력자배제 (해당질환 및 시술력없는자)	Aortocoronary artery bypass graft		CABG 처치코드	당뇨병진단전
9		Angina pectoris	I20		당뇨병진단전
10		Cerebral infarction	I63		당뇨병진단전
11		Transient ischemic attack	G458, G459		당뇨병진단전
12	과거 이상지혈증 약물복용자배제	All Statin		약물코드	당뇨병진단전
		fibrate		약물코드	당뇨병진단전
13	심한 고혈당자 배제	Hba1c < 12.0%		검사코드 및 수치	모든 시간

각 데이터베이스에서 연구대상자를 선별 및 비교하기 위해 대상자 관련변수를 정리하였다. 각각은 해당요인, 관련코드(KCD 코드, 약물코드, 검사 및 처치코드 (EDI) 코드), 추가정보 등으로 정리하였다.

2) 혼란변수

동일하게 혼란변수에 대해서도 다음과 같이 정리하였다.

<표 3-3> 혼란 변수 및 관련내용

용도	요인	KCD 코드	기타코드	추가정보
1	성			진단시
2	연령			진단시
3	체질량지수	Body Mass index		
4	고혈압여부	고혈압진단(진료)	I10,I11,12,13,15	당뇨병진단전
		고혈압약제	고혈압약물코드	1일 이상 처방.
5	흡연여부			당뇨병진단전
6	당뇨병성 망막병증	Diabetic retinopathy 진단	E113	당뇨병진단후 1년내
		Diabetic nephropathy진단	E112	당뇨병진단후 1년내
7	당뇨병성 신병증	Spot urine A/C ratio	검사코드 및 수치	당뇨병진단후 1년내
		GFR < 60	검사코드 및 수치	당뇨병진단후 1년내
8	콜레스테롤수치	HDL-cholesterol	검사코드 및 수치	당뇨병/약물후18개월
		LDL-cholesterol	검사코드 및 수치	당뇨병/약물후18개월

혼란변수로서 고혈압여부는 고혈압진단 및 1일 이상의 고혈압약제처방으로 정의하였다. 당뇨병성 망막병증은 당뇨병진단후 1년이내의 진단코드로 정의하였으며, 당뇨병성 신병증은 당뇨병진단후 1년이내의 진단코드 또는 소변내 Albumin-Creatinin ratio 또는 사구체여과율 60미만으로 정의하였다. LDL-C와 HDL-C도 혼란변수로 고려하였다.

3) 추적 기간

CDM 5.3.0_02 데이터는 2020년말까지 추출된 데이터를 바탕으로 하여 분석하였다. DW를 바탕으로 연계된 공단자료는 2020년 12월 13일까지의 자료로 연구는 2020년말까지 추적되었다.

결과변수에 대해서 Statin 사용군에서 약물사용에 의한 결과로 포함되기 위한 시간으로 3년을 lag period로 잡았다.

4) 결과변수

첫째, 콜레스테롤수치 감소율 및 조절율을 LDL-cholesterol(LDL-C) 및 non-HDL-cholesterol(non HDL-C)로부터 확인하고자 한다.

둘째, 심혈관질환발생을 보고자 한다. 급성심근경색, 관상재관술 시술 및 수술(관상동맥재관류 시술 및 CABG수술시행), 급성뇌경색, 급성뇌출혈

셋째, 모든 원인사망위험 위험에 대해서 비교하였다.

3. 각 데이터베이스의 종류와 활용

1) 일산병원 DW(구) 및 해당환자의 공단연계자료

DW(구)는 비교대상이자 분석의 기초가 되는 데이터 베이스이다. CDM 데이터 베이스 및 CDW 데이터베이스의 현황을 분석함에 있어 활용한 이번 연구에서 공단연계를 위한 기본적인 대상자를 설정을 일산병원 DW를 통해서 시행한다. 또한 DW바탕으로 Sybase 프로그램으로 추출된 대상자에 공단자료를 연계하여 2002년부터 2020년까지의 공단자료 및 청구자료를 확보하였으며 이를 바탕으로 과거력 배제, statin제제 약물사용여부를 확인하는데 사용되었다. 또한 최종분석을 통해 결과변수 확인에 활용하였다.

또한, Sybase를 통해 대상자 선정을 위한 다양한 조건별로 인원수를 구하였으며 이를 CDW데이터 베이스로부터 구한 인원수로 비교하여 현재 CDW데이터베이스의 질의 파악하는데 활용하였다.

2) 일산병원 CDM 5.3.0 02, weekly 데이터베이스

CDM 데이터 베이스의 현황을 파악을 위한 매핑틀 등의 데이터질 분석에 활용하였다. 또한 CDM 자체 분석도구인 ATLAS를 활용하여 연구방법에 맞는 대상자를 수립하고 이를 바탕으로 추정(Estimation)을 위한 분석코드 생성에 이용하고자 한다. 나아가 이를 바탕으로 Feedernet platform활용하여 두 군간의 PS matching을 포함한 Estimation 실행 진행하고자 하였다.

3) 일산병원 CDW 데이터베이스

새롭게 구축된 CDW 데이터베이스로 DW로부터 기존 데이터를 이관되어 왔다. 따라서 기존 DW데이터베이스와의 비교를 위해 여러 조건을 통해서 CDW데이터베이스로부터 MSSQL 프로그램을 통해 인원수를 구하였으며 이를 바탕으로 DW에서 산출한 인원수와 비교하여 CDW 데이터질검증을 시행하였다. 또한 CDW자료가 웹 인터페이스인 iDEA로 잘 구현되는지 확인하기위해서 iDEA를 통해서 대상자조건에 따른 인원수를 산출하였으며, 이를 CDW로부터 직접 산출한 인원수 및 DW로부터 산출한 인원수와 비교 확인하였다.

4. 분석방법

CDM에서는 자체탑재된 분석도구인 ATLAS를 활용하여 Estimation(추정)을 위한 분석코드를 생성하였으며, 이를 바탕으로 Feedernet 플랫폼에서 Propensity score matching을 포함한 Cox regression을 시행하였다.

DW를 기반으로 한 분석에서는 SAS 9.4를 활용하여 연구대상자의 특성을 분석하였으며 Cox regression을 시행하였다.

제4장

연구 결과

제1절 국민건강보험 일산병원 CDM 현황	47
제2절 국민건강보험 일산병원 CDW 현황	52
제3절 당뇨병환자에서 Statin제제 사용에 따른 LDL-C와 심혈관질환 위험 및 사망위험 분석	56

제4장 연구 결과

제1절 국민건강보험 일산병원 CDM 현황

1. 데이터의 질

1) 데이터 베이스내 코드 및 레코드 매핑 현황

<표 4-1> 국민건강보험 일산병원 CDM 5.3.0_02버전 매핑률

(2021. 12. 31. 현재)

도메인	원본 레코드수	매핑된 레코드수	매핑 레코드 비율 (%)	원본 코드수	매핑된 코드수	매핑코드 비율(%)
Person	1,367,171	1,367,171	100.0%			
Condition_occurrence (진단)	26,432,955	26,307,899	99.5%	15,592	14,328	91.9%
Drug_exposure (약물)	83,237,465	83,027,507	99.7%	5,143	4,932	95.9%
Measurement (검사)	290,183,181	278,217,168	95.9%	2,984	2,204	73.9%
Procedure_occurrence (시술처치)	103,313,825	86,622,017	83.8%	16,831	14,306	85.0%
Visit_occurrence (진료방문)	16,577,348	16,576,618	100.0%	4	4	100.0%
Observation_period (추적기간)	1,367,171	1,367,171	100.0%			
Specimen (검체)	22,922,827	22,706,054	99.1%	403	357	88.6%
Observation (조사정보)	7,605,094	5,963,700	78.4%	308,347	308,284	100.0%
Death	21,130	21,130	100.0%			

일산병원의 CDM 02버전의 매핑 상태를 확인한 결과 총인원 1,367,171명에 대해서 데이터베이스에 레코드(기록값)를 가지고 있는 값들 중에서 진단, 약물, 검사 등은 모두 95% 이상의 매핑률을 보이고 있었다. 수술 시술 처치에 해당되는 Procedure에 대한 부분 84%, 검사결과에 해당되는 Measurement 도메인은 73.9%, 주호소증상, 흡연, 음주 등 조사 항목에 해당되는 observation 의 경우 78% 정도의 매핑률을 보였다.

이는 시기적으로 다르기는 하지만 21년 7월 분산형 바이오헬스빅데이터 사업단에서 발간한 공통데이터모델품질 검사 보고서에서 아주대병원 CDM 데이터베이스에 대해 시행된 품질 검사상 제시된 Procedure occurrence 도메인의 23.5%, observation의 67.9%에 비해서는 매핑률이 높은 수치이다.

<표 4-2> 아주대병원 CDM 5.3.0 버전 매핑률¹²⁾

(2021. 12. 31. 현재)

도메인	원본 레코드수	매핑된 레코드수	매핑 레코드 비율 (%)
Person	2,714,449	2,714,449	100.0%
Condition_occurrence (진단)	37,602,513	7,341,428	99.3%
Drug_exposure (약물)	96,736,317	91,209,388	97.2%
Measurement (검사)	448,277,322	440,814,934	98.3%
Procedure_occurrence (시술처치)	367,104,816	86,254,749	23.5%
Visit_occurrence (진료방문)	28,039,978	28,039,978	100.0%
Device_exposure (기기)	196,736,317	191,209,388	97.2%
Observation (조사정보)	23,118,179	15,699,752	67.9%
Death	26,299	26,299	100.0%

반면에 일산병원에서 새롭게 구축된 CDW기반으로 2022년 1월부터 사용가능해진 5.3.1 weekly 버전의 경우에 Measurement와 Procedure occurrence 가 각 32.6% 39.3%로 낮은 매핑률을 보이고 있는 상태이다. <표 4-3>

12) 공통데이터모델 품질 검사 보고서: 아주대학교병원, 2021

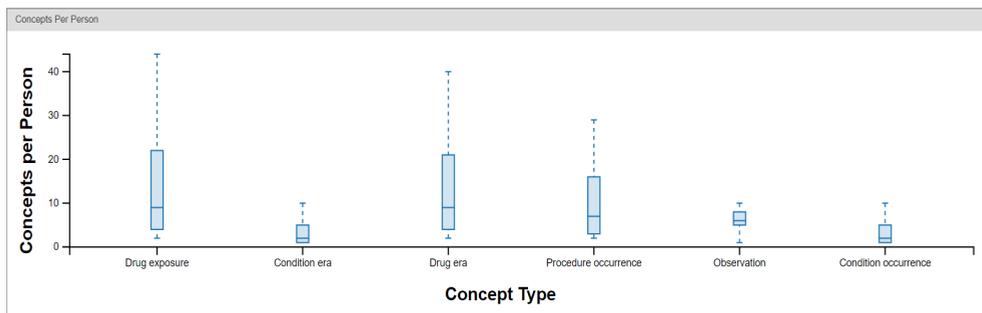
48 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

<표 4-3> 국민건강보험 일산병원 CDM 5.3.1_weekly 버전 매핑률

(2022. 2. 15. 현재)

도메인	원본 레코드수	매핑된 레코드수	매핑 레코드 비율 (%)
Person	1,437,911	1,437,911	100.0%
Condition_occurrence (진단)	80,863,712	80,844,442	100.0%
Drug_exposure (약물)	165,579,156	165,303,585	99.8%
Measurement(검사)	610,480,259	198,809,834	32.6%
Procedure_occurrence (시술처치)	483,807,958	190,104,633	39.3%
Visit_occurrence (진료방문)	19,037,704	18,937,402	99.5%
Device_exposure(기기)	43,111,333	33,580,291	77.9%
Observation(조사정보)	21,518,500	2,837,286	13.2%
Death	23,664	23,664	100.0%

2) 인원당 매핑된 Concept ID수



[그림 4-1] 국민건강보험 일산병원 CDM 5.3.0_02버전 인원당 각 테이블(도메인)의 Concept ID수

또 다른 데이터의 질을 평가하는 방법으로 인원당 매핑된 Concept ID수가 있다. Concept ID는 OMOP-CDM에서 모든 분석에 기반되는 모든 진단, 검사, 처치 등에 대한 기본적인 표현이다. 상급종합병원 및 종합병원의 진료 형태가 크게 다르지 않은 한국의 상황에서 모든 데이터베이스의 레코드수의 매핑률 만큼이나 인원별 Concept ID수를 다른 병원과 비교할 때 데이터의 질을 확인할 수 있다. 데이터의 질을 평가하는 CDM 자체 도구인 Achilles를 활용하여 인원당 매핑된 concept ID 수를 확인하였다.

약물사용의 도메인인 Drug exposure의 중위값은 9.0, 진단에 해당되는 Condition occurrence의 중위값은 2.0, Procedure 7.0, observation 6.0으로 나타났으며, 이는 2017년 16개 기관의 CDM 데이터베이스를 기준으로한 시행한 품질검사에 비해서 떨어지지 않는 수치였다²².

2. 주요 연구 변수별 매핑 현황

이번 연구에서 사용하는 변수에 대해서 원내처방코드와 이에 해당되는 레코드가 매핑된 표준용어에 대한 OMOP-CDM에서의 Concept ID를 확인하였다.

<표 4-4> 국민건강보험 일산병원 CDM 5.3.0_02버전 진단검사의학적 변수의 매핑률

변수	원내 처방 코드	매핑된 표준용어 concept_ID	매핑률
LDL-Cholesterol	L303001	원내 처방내역 없음	100%
	L3062	3028437	
	L306201	3028437	
	L3062D	3028437	
	L3062F	원내 처방내역 없음	
Triglyceride	L3029	3022192	100%
	L3029D	3022192	
	L3029F	원내 처방내역 없음	
	L8050	3022192	
Glycated hemoglobin(Hba1c)	L3034	3004410	100%
	L30341	원내 처방내역 없음	
	L3034F	원내 처방내역 없음	
Albumin Craetinine ratio(A/C ratio)	L312701	3034485	100%
	L812701	원내 처방내역 없음	
	GL3020	원내 검사내역 없음	
Glomerular filtration rate(GFR)	GL8020	원내 검사내역 없음	100%
	L30201F	46236952	
	L30202F	46236952	
	L30203F	40764999	
	L80201F	46236952	
	L80202F	46236952	
	L80203F	40764999	
Total Cholesterol	L3013	3027114	100%
	L3013D	3027114	
	L3013F	원내 처방내역 없음	
	L8049	3027114	
HDL-cholesterol	L3030	3007070	100%
	L3734	원내 처방내역 없음	
	L3030D	3007070	
	L3733	원내 처방내역 없음	
	L3735	원내 처방내역 없음	

<표 4-4> 에서 보는 것처럼 이번 연구에서 사용되는 콜레스테롤 관련 검사들과 Hba1c, 소변내 Albumin-Creatinine ratio, 사구체 여과율을 살펴보았을 때, 여러개의 검사항목들에 대해서 존재하는 레코드에 대해서는 모두 100%의 매핑률로 CDM 데이터 베이스에 매핑되어 있었다.

또한, 관상동맥 재관류를 위한 PCI(Percutaneous Coronary Intervension)와 CABG (Coronary Artery Bypass Graft)에 해당되는 원내처방코드와 이에 매칭되는 건강보험 청구코드인 EDI 코드를 확인하였으며, 각각 코드에 대해서 OMOP-CDM concept ID가 잘 매칭되고 있음을 확인하였다. (부록1)

3. CDM 데이터베이스내 주요 혼란변수 및 대상자 비교

CDM 자체로 제공하는 분석도구인 ATLAS내에서는 각 도메인테이블의 상태 및 각 도메인내의 Concept별 상태를 확인할수 있는 Data sources 세션을 제공하고 있다. 이를 통해서 연구에서 주되게 사용가능한 혼란변수에 대해서 확인하였다. 우선 체질량지수의 경우, NHIMC CDM 5.3.0_02 데이터 베이스에 대해서 486,944명으로 전체인원수의 35.6%가 체질량지수에 대한 수치를 가지고 있었으며 평균적으로 한명당 3.79개의 레코드를 가지고 있었다.

하지만, 심혈관질환의 주요 혼란변수로 사용되는 흡연에 대해서는 Observation 도메인에 해당되는 것으로 흡연에 대한 레코드수는 전체 인원의 5%정도 밖에는 되지 않아 연구시행에 있어 주요 혼란변수로 사용하기에는 어려움이 있을 것으로 판단된다.

<표 4-5> NHIMC 5.3.0_02버전 데이터베이스내 흡연관련 Concept 현황

Concept Name	Concept ID	Person count	Prevalence
Current smoker	40766945	2,112	0.15%
Ex-smoker	4310250	3,174	0.23%
Non-smoker	4222303	53,620	3.92%
Smoker	4298794	2,101	0.15%
Stopped smoking	4052032	1,947	0.14%
Tobacco smoking consumption	4144271	948	0.07%
Tobacco user	4005823	131	0.01%

또한, 실제 현재 사용가능한 CDM 데이터 베이스인 NHIMC 5.3.0_02 버전과 5.3.1 weekly 버전에서 각각 ATLAS를 활용하여 <표 3-2> 에 맞춰서 대상자를 선정하였을

때, NHIMC 5.3.0_02 버전에서는 20,695명이 산출되었으나 5.3.1 weekly 버전에서는 16,103명이 산출되어 현재로서는 5.3.1 weekly 버전의 경우 매핑률이 떨어지면서 연구에의 활용이 부적절할 것으로 판단된다.

제2절 국민건강보험 일산병원 CDW 현황

1. 구DW와 CDW 통한 산출 대상자수 비교확인

<표 4-6> 구DW, CDW 및 iDEA에서 산출된 대상자수 비교

	Definition	factor	KCD7 code	추가정보	CDW 인원수	i-DEA 인원수	DW(구) 인원수
1	Type 2 DM	diabetes	E11-14	CDW상 첫 진단을 구하기 힘들			
2	2002-2015	진단일자			39,414	39,414	39,414
3	Age 30-79	진단당시 나이		진단시 나이를 case design에서 구하게 되면 실제 초진일 누락 가능성			
		anti-diabetic drugs		첫 진단일 이후 10년 이내 (3650일) 진단받은 내역	33,669	33,669	33,669
4	기저콜레스테롤 수치	LDL-cholesterol >70mg/dl		당뇨병 진단전 6개월 내에 LDL-C ≤ 70mg/dl 또는 당뇨병 진단전 6개월 내에	1,767	1,767	1,763
5		Triglyceride <500mg/dl		500mg/dl ≤ TG ≤ 2000 mg/dl 인 환자 제외	391	391	391
6	과거력자 배제	myocardial infarction	I21, I22	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역	4,433	4,433	4,433
7		angina pectoris	I20	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역			
8		cerebral infarction	I63	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역			
9		Transient ischemic attack	G458 G459	당뇨병 첫 진단 이전 10년까지의 기간 중 진단받은 내역			
10	과거 이상지혈증약물 복용자 배제	placement of coronary stent		당뇨병 첫 진단 이전 10년까지의 기간 중 처방받은 내역	792	792	792
11		Aortocoronary artery bypass graft		당뇨병 첫 진단 이전 10년까지의 기간 중 처방받은 내역			
12	고혈당자 배제	all Statin		당뇨병 첫 진단 이전 2년까지의 기간 중 처방받은 내역	6,816	6,816	6,816
		fibrate		당뇨병 첫 진단 이전 2년까지의 기간 중 처방받은 내역			
13	고혈당자 배제	Hba1c < 12%		모든 시간	2,624	2,624	2,624

52 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

국민건강보험 일산병원 CDW 데이터베이스의 현황을 파악하기 위해 구DW로부터 <표 3-2>를 통해 정리된 대상자조건에 대해서 Sybase 프로그램을 통해서 각 조건에 맞는 인원수를 산출하였다. 또한, 동일한 조건들에 대해서 CDW 데이터 베이스로부터 MSSQL프로그램을 통해서 인원수를 산출하였다. 마지막으로 CDW의 웹인터페이스에 해당되는 iDEA를 이용하여 각 조건에 대한 인원수를 산출하여 각각을 비교하였다. 당뇨병진단코드, 당뇨병진단시 나이 및 일자를 바탕으로한 인원수는 모든 경우에 39,414명으로 동일하였다.

또한 혈액검사 수치에 해당되는 LDL-C, Triglyceride, Hba1c 와 관련된 조건에서도 모두 동일한 인원수를 보여주었다. 또한 과거력을 통해 대상자를 배제하기 위해 과거심혈관질환 진단명 및 관상동맥재관류와 관련된 PCI, CABG 시행에 대한 대상자에서도 세 경우에 모두 동일한 인원수가 산출되었다.

이를 바탕으로 현재 새롭게 구축된 CDW 데이터베이스에는 구DW 시스템의 데이터가 상당수준에서 잘 이관되었음을 확인하였다. 또한, CDW의 웹인터페이스인 iDEA를 통해서 CDW의 대상자 산출도 잘되고 있음을 확인할 수 있었다.

2. 웹인터페이스 iDEA의 기능 현황 및 기능적 제한점

실제 연구자들이 CDW를 활용하여 연구를 진행시에 실제적으로 이용하게 되는 iDEA의 기능 현황을 확인하였다.

1) 여러 진단명을 대상자로 그룹화할시, 특정개인을 기준으로 선택기능 부재

이번 연구에서 iDEA를 이용하여 당뇨병에 해당되는 KCD 코드인 E11,12,13,14를 가지고 대상자를 산출할 때, KCD코드의 세부적인 진단명에 대해서 각각의 대상자가 산출되는 현상이 나타났다.

예를 들어 <표 4-7> 와 같이 KCD E11 코드의 하부코드인 E119에 대해서도 13개의 진단코드로 세부적으로 더 분류가 되며 각각 13개의 진단명에 대해서 별도의 진단코드 그룹이 만들어지는 상황이다. 특히 특정개인에 대해서 E119의 동일한 코드라 하더라도 병원내에서 각각의 세부 진단코드로 다른 날짜에 진단명이 기입된 경우에 E119를 대상으로 대상자를 산출할 때 특정개인이 여러번 중복되는 결과가 발생한다. 특정개인을 기준으로 가장 빠른 날짜의 진입을 선택하는 등의 선택기능이 부재함으로써 발생하는 현상이다. 이를 수정하기 위해서는 1차적으로 진단명에 해당되는 인원을 모두 다운로드한 후에 중복을 제거하여 업로드후에 이후 작업을 진행해야 한다.

<표 4-7> KCD7 E119 코드에 대한 세부사항 예

KCD	진단코드	진단명
E119	105083	Type 2 DM
E119	202610	Maturity-onset diabetes (mellitus, nonobese, obese), without complications
E119	202613	Maturity-onset diabetes (mellitus, nonobese, obese), without complications
E119	202619	Non-insulin-dependent diabetes of the young, without complications
E119	202611	Nonketotic diabetes (mellitus, nonobese, obese), without complications
E119	202617	Stable diabetes(mellitus, nonobese, obese), without complications
E119	105084	Type 2 DM without complications
E119	105085	합병증을 동반하지 않은 2형 당뇨병
E119	202612	합병증을 동반하지 않은 비케톤성 당뇨병(진성, 비비만성, 비만성)
E119	202614	합병증을 동반하지 않은 성숙기발병당뇨병(진성, 비비만성, 비만성)
E119	202616	합병증을 동반하지 않은 성인발병당뇨병(진성, 비비만성, 비만성)
E119	202618	합병증을 동반하지 않은 안정성 당뇨병(진성, 비비만성, 비만성)
E119	202620	합병증을 동반하지 않은 청년의 인슐린비의존 당뇨병

2) 조건그룹이 5개초과시 자동으로 괄호() 생성후 선실행 현상

iDEA는 매우 사용자친화적인 프로그램으로 연구자가 대상자를 조건을 걸어 산출할 때 기본조건이 되는 대상자에 추가적인 조건들을 간단하게 옆으로 붙여서 원하는 조건에 맞는 대상자로 그 범위를 줄여나갈 수 있다. 그런데 프로그램의 기능상 6개이상의 조건에 대해서는 마지막 두 개의 조건에 대해서 같은 그룹으로 괄호()가 설정되어 괄호()가 선실행후에 앞의 4개의 조건에 붙게 된다. 예를 들어 연구자는 [그림4-2]에서 앞의 4개의 조건후에 과거 CABG, PCI를 시행한 대상자를 빼고, 그 이후에 과거 Statin이나 fibrate 제제를 복용한 사람을 또 제거하고 싶는데, 이 두 개의 조건이 먼저 선실행([그림4-2]에서 보라색의 두 개 조건) 되면 과거 CABG, PCI를 시행한 사람중에 statin, fibrate 제제를 복용했던 사람을 제거한 후에 이 대상자를 앞의 4개 조건에서 남은 대상자로부터 제거하여 마지막 대상자를 산출하는 오류가 발생하게 되는 것이다. 따라서, 이를 예방하기 위해서는 여러개의 조건에 이용될 경우에는 다운로드와 재업로드를 활용해야 할 필요가 있다.



[그림 4-2] iDEA 내의 조건그룹 실행오류 예

54 당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

3. CDW 데이터 베이스의 혼란변수 현황

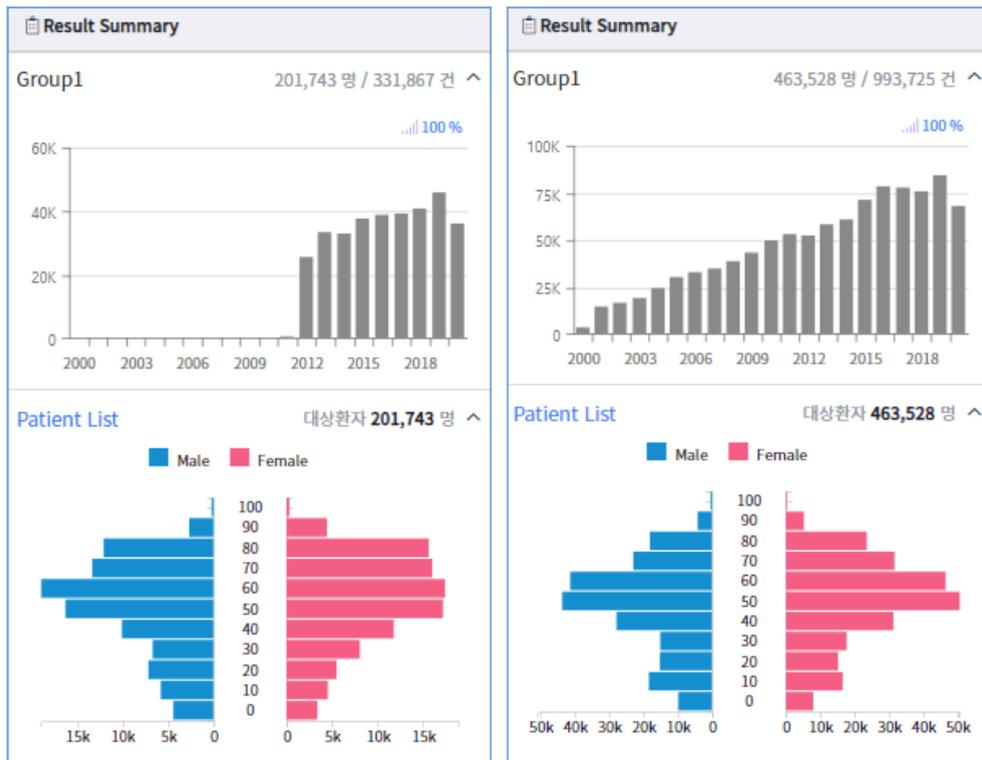
CDW 데이터베이스에서 실제 심혈관질환에 대한 연구를 진행할 때 사용할수 있는 변수인 체질량지수와 흡연에 대해서 확인하였다.

흡연의 경우 CDM 5.3.0_02에 비해서는 많은 201,743명의 자료를 가지고 있었다. 다만, 이는 전자의무기록으로부터 추출된 자료이므로 전자의무기록 시스템에 구축되기 전의 2011년 이전의 자료는 부재했다. 이는 추후 다른 심혈관질환연구에 있어서 본원의 흡연관련 변수를 2011년이전에 대해서 사용이 어렵다는 것을 의미하며, 전자의무기록으로부터 추출되는 증상이나 주호소 등의 다른 변수에 대해서도 동일한 현황일수 있음을 의미한다.

체질량지수에 대해서 확인한 결과 463,528명의 데이터를 가지고 있었다. [그림 4-3]

a. 흡연여부

b. 체질량 지수



[그림 4-3] CDW 데이터베이스내 흡연 및 체질량지수 데이터 현황

제3절 당뇨병환자에서 Statin제제 사용에 따른 LDL-C와 심혈관 질환 위험 및 사망위험 분석

1. CDM 자료를 활용한 non-matching 및 PS matching

당뇨병진단후 2년내 시작된 statin 처방이 1년이상 유지된 군을 statin 사용군으로, 당뇨병 진단후 2년동안 statin 처방이 없었던 군을 statin 비사용군으로 설정하였다. CDM 데이터베이스에서는 PS 매칭전에는 statin 사용군이 1,891명, statin 비사용군이 7,982명이었으며, PS 매칭 후에는 두군 모두 1,809명이었다. PS 매칭전에는 statin 비사용군에 비해서 Statin사용군에서 모두 심근경색 발생위험은 2.02배, 관상동맥재관류는 3.55배, 급성뇌경색은 1.55배 유의미하게 상승한 것으로 나타났다. PS 매칭후에는 모두 심근경색 발생위험은 2.0배, 관상동맥재관류는 1.5배, 급성뇌경색은 1.68배 유의미하게 상승한 것으로 나타났다. <표 4-8>

하지만 모든 원인 사망에 대해서는 Statin 비사용군에 비해서 Statin 군에서 PS매칭전에 0.74배 PS 매칭후 HR 0.56배으로 유의미하게 사망위험이 감소했다.

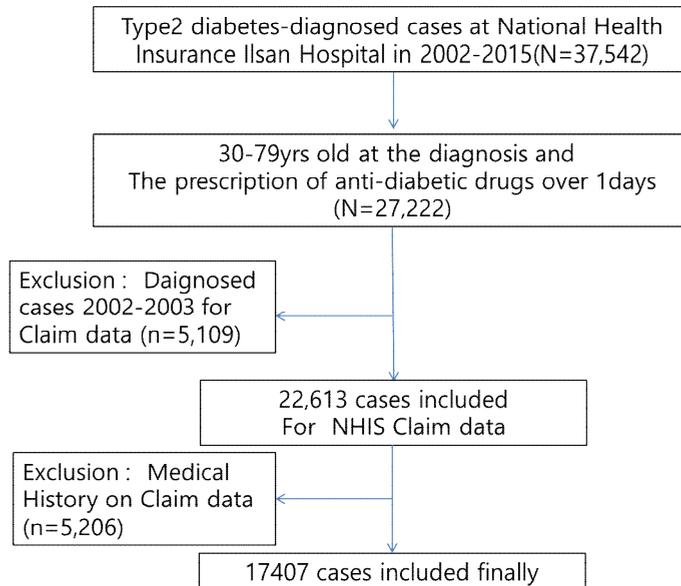
<표 4-8> CDM데이터베이스를 활용한 당뇨병환자에서 Statin사용여부에 따른 심혈은 질환 및 사망위험(PS 매칭전)

	non-Statin users(N=7932)			Statin users(=1891)			
	Event n(%)	Person years	IRate per 1000py	Event n(%)	person year	IRate per 1000py	Hazard Ratio(95%CI)
primary Outcome							
Acute Myocardial infarction	61(0.7)	35,440	1.72	49(2.5)	13,751	3.56	2.02 (1.38-2.95)
Coronary revascularisation	5(0.06)	35,471	<0.14	5(0.2)	13,915	<0.36	3.55 (0.78-18.12)
Ischemic Stroke	119(1.5)	35,342	3.37	72(3.8)	14655	5.33	1.55(1.15-2.08)
hemorrhagic Stroke	40(0.5)	35,142	1.14	19(1.0)	13,861	1.37	1.11(0.62-1.89)
Secondary outcome							
All cause Death	387 (2.51)	35,473	10.91	122 (6.4)	13,981	8.73	0.74 (0.60-0.91)

<표 4-8> CDM데이터베이스를 활용한 당뇨병환자에서 Statin사용여부에 따른 심혈관질환 및 사망 위험(PS 매칭후)

	non-Statin users(N=1809)			Statin users(=1809)			
	Event n(%)	Person years	IRate per 1000py	Event n(%)	person year	IRate per 1000py	Hazard Ratio(95%CI)
primary Outcome							
Acute Myocardial infarction	15(0.8)	7,936	1.89	49(2.7)	13,742	3.57	2.00 (1.05-4.02)
Coronary revascularisation	5(0.2)	7,974	<0.63	5(0.2)	13,905	<0.36	1.50 (0.25-11.29)
Ischemic Stroke	32(1.7)	7,910	4.05	72(3.9)	13,497	5.33	1.68 (1.03-2.79)
hemorrhagic Stroke	8(0.4)	7,933	1.01	19(1.0)	13,851	1.37	1.29 (0.48-3.60)
Secondary outcome							
All cause Death	81(4.5)	7,993	10.13	122 (6.7)	13,971	8.73	0.56 (0.37-0.83)

2. DW기반 공단연계자를 활용한 연구분석



[그림 4-4] 연구대상자 흐름도

DW를 기반으로 한 공단연계자료 분석에서 시행한 분석은 다음과 같다.

2002년부터 2015년까지 본원에서 진단받은 당뇨병환자는 37,542명이었다. 그러나, 2002년부터 청구자료가 존재하는 공단자료의 특성상 2년치이상의 과거력을 확보하기 위해서 2002년부터 2003년까지 진단받은 당뇨병환자 5,109명을 배제하기로 하였다. 이에 2004년부터 2015년까지 국민건강보험 일산병원에서 진단받은 30-79세의 당뇨병환자 22,613명에 대해서 2002년부터 2020년까지의 공단자료를 연계하였다.

그 중 과거 심혈관질환 등의 과거력, statin제제를 복용력등을 제외하고 17,407명에 대해서 최종적으로 분석을 시행하였다. 당뇨병 진단 2년이내 1년이상 연속적으로 statin을 먹은 1,809명, Statin제제를 당뇨병 진단후 2년동안 처방이 없었던 비사용군 1,414명을 대상으로 분석하였다.

대상자 기본특성상 Statin사용군은 평균 55.9세, Statin 비사용군은 57.8세였다. Statin 비사용군이 남자가 더 많았고 당뇨병진단시 혈압은 모두 두군에서 42%정도 보유하고 있었다. 기저 LDL-C ldl수치가 Statin사용군은 평균 134mg/dl였고 비사용군은 106mg/dl였다.

<표 4-9> 연구대상자 기본특성

	Non-Statins User N=1414	Statins User N=1809
Age, years, Mean±SD	57.8±12.4	55.9±11
30-39 years, n(%)	117(8.3)	123(6.8)
40-59 years, n(%)	644(45.5)	996(55.1)
60-79 years, n(%)	653(46.2)	690(38.1)
Sex		
Male, n(%)	932(65.9)	999(52.2)
Female, n(%)	482(34.1)	810(44.8)
Body mass index, kg/m ²	24.4±4.9	25.4±4
Hypertention, n(%)	591(41.8)	765(42.3)
Diabetic nephropathy, n(%)	502(35.5)	649(35.8)
Diabetic retinopathy, n(%)	52(3.7)	113(6.3)
lipid, mg/dl, Mean±SD		
Total chol	173.4±40.1	211.9±42
Triglyceride	146.6±88.1	181.6±122.6
HDL-C	48.7±12.5	49.2±12.3
LDL-C	105.9±22.8	134.4±30.9
Hba1c,	7.5±1.9	7.7±1.8

연속적으로 18개월 이상 먹은 statin제제를 복용한 군을 대상으로 LDL-C 및 non-HDL-C 수치 및 조절율 및 감소율을 확인하였다. Statin 비사용군은 당뇨병 진단후 18개월이상 statin제제를 처방받지 않은 군으로 하였다. 조절율은 LDL-C을 100mg/dl 미만으로 조절되는 비율을, non_HDL-C 은 130mg/dl 미만으로 조절되는 정도를 확인하였다. 감소율은 기저치로부터 18개월이후의 수치의 변화를 확인하였다.

Statin 사용군에서 처음 사용으로부터 18개월후 LDL-C, non-HDL-C 수치는 각각 목표치에 대해서 61%, 63% 조절되고 있었다. 이는 statin 비사용군의 52%, 58%에 비해 높은 수치였다. 또한, Stain 사용군에서 기저치에 비해서 LDL-C 27.9%정도 감소한 수치를 보였다.<표 4-10>

<표 4-10> Statin 사용 18개월후 콜레스테롤 수치 변화

	N	LDL-C		Control rate for <100mg/dl	% Change of LDL-C from Base	Non-HDL		Control rate for <130mg/dl	% Change of non-HDL-C from Base
		base	>18m			Base	>18m		
Non-Statin	1414	105.5	99.3	52%	-4.3%	132	124.6	58%	-3.2
Statin	1809	134.4	94.3	61%	-27.9%	167.2	120.3	63%	-28.3

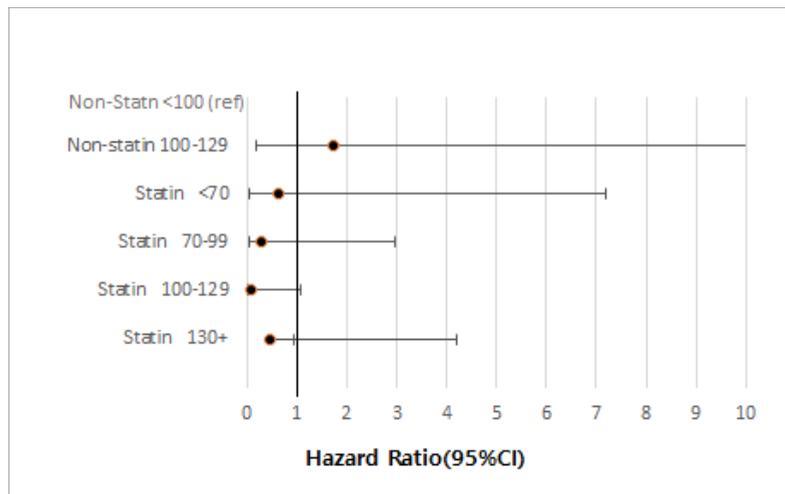
Statin 사용군과 비사용군간의 심혈관질환의 발생 및 모든 원인에 의한 사망위험을 비교하였다.

<표 4-11> DW기반 공단연계자료에서 Statin사용군과 비사용군간의 심혈관질환 발생 및 사망위험

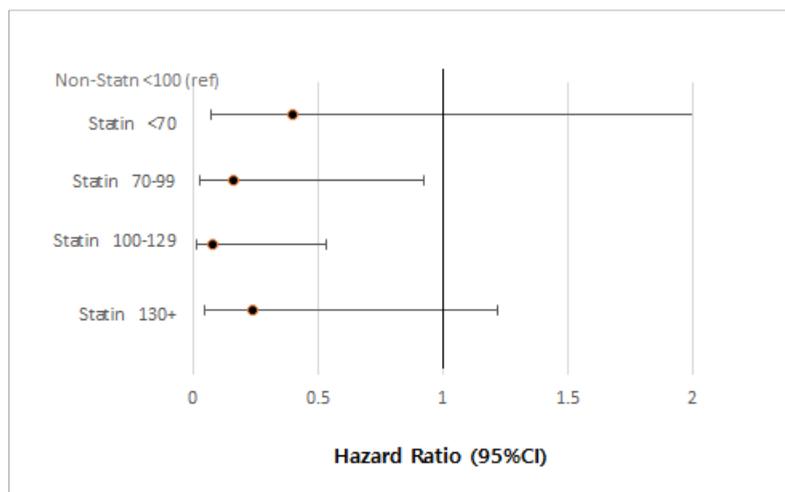
	non-Statin users(N=1414)				Statin users(=1809)			
	Event n(%)	Person years	IRate per 1000py		Event n(%)	person year	IRate per 1000py	Hazard Ratio(95%CI)*
primary Outcome								
Acute Myocardial infarction	7(0.3)	13207	0.53		39(2.1)	14655	2.66	0.86 (0.358-2.082)
Coronary revascularization	7(0.3)	13210	0.53		69(3.8)	14221	4.85	0.75 (0.339-1.654)
Ischemic Stroke	1(0.1)	13303	0.07		11(0.6)	15094	0.72	0
hemorrhagic Stroke	2(0.1)	13305	0.15		2(0.1)	15129	0.13	0
Secondary outcome								
All cause Death	259 (18.3)	10316	25.1		176 (9.6)	14731	11.94	0.52 (0.419-0.640)

연령 성별 보정을 했을때 스타틴군에서 급성심근경색, 관상동맥 재관류술 발생위험이 CDM 데이터베이스 결과와는 반대로 Statin사용군에서 발생위험이 각각 0.86배, 0.75배로 감소하였으나, 통계적으로 유의하지는 않았다. 다른 혼란변수를 모두 보정한 경우에는 통계적인 시행이 불가능했다. 모든 원인의 사망에 대해서는 비교한 결과, statin 사용군에서 0.52배로 비사용군에 비해서 48%의 사망위험이 감소하는 것으로 나타났다.<표 4-11>

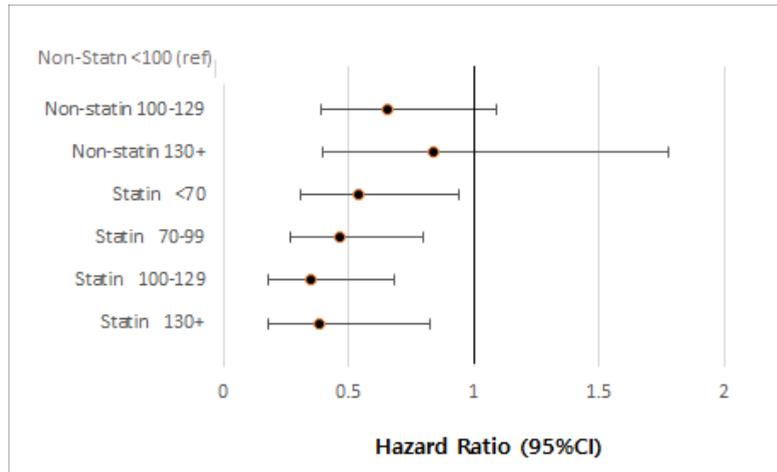
a. 급성심근경색



b. 관상동맥 재관류술



C. 모든 원인 사망



[그림 4-5] DW기반 공단연계자료에서 Statin사용군과 비사용군간의 LDL-C에 따른 심혈관질환 발생 및 사망위험

마지막으로 Statin 비사용군에서 LDL-C < 100 mg/dL 인 그룹을 참고군로하여 statin 사용군에서 statin사용후 18개월 이후의 수치로 LDL-C 70mg/dl 미만인 그룹, 70-99mg/dl 인 그룹, 100-129mg/dl 인 그룹, 130이상인 그룹에서의 심혈관질환 발생위험 및 사망위험을 분석하였다.

LDL-C 100mg/dl 미만인 statin 비사용군과 비교할 때 statin 사용군에서 모든 LDL-C 수치별 그룹에서 급성심근경색과 관상동맥재관류술 시행의 발생위험이 낮았으나 통계적으로 유의하지는 않았다. 다만, 모든 원인의 사망에 대해서 LDL-C 100mg/dl 미만인 Statin 비사용군에 비해 Statin 사용의 세 LDL-C그룹 모두에서 사망위험이 유의하게 감소하였다.

제5장

고찰 및 제언

제1절 결론 및 고찰	65
제2절 제언	68

제5장 고찰 및 제언

제1절 결론 및 고찰

우리는 본 연구를 통해서 당뇨병환자에서 statin제제의 사용여부에 따른 심혈관질환의 발생위험 및 사망위험에 대한 분석을 통해서 국민건강보험 일산병원에 구축된 CDM 데이터베이스 및 CDW 시스템과 웹인터페이스인 iDEA를 현황을 확인하였고, DW자료를 기반으로 공단자료에 연계하여 최종적인 분석을 시행하였다.

현재 일산병원 CDM인 NHIMC 5.3.0_02 버전은 모든 테이블(도메인)에서 70%이상의 매핑률을 보이고 있으며, 특히 진단과 관련된 Condition Occurrence, 약물에 대한 Drug exposure 의 경우에 모두 90%이상의 매핑률을 보여주고 있으며, 대규모의 130만명의 데이터를 활용가능하다. 또한, 각 주된 연구변수로 사용되는 진단검사적인 변수에 대해서 매핑률도 높은 상태이다. 자체 ATLAS의 탑재된 기능과 Feedernet플랫폼을 통해서 PS matching을 포함한 estimation, prediction 시행가능한 상태이다. 다만 아직까지 smoking 과 같은 Observation에 해당되는 정보가 제한적이므로 이를 이용한 연구는 제한이 있을 것으로 판단된다. 최근 CDW를 기반으로 한 CDM 5.3.1 weekly 버전의 경우 데이터소스를 CDW, DW로 병행중이며 지속적으로 업데이트 중이나 02버전에 비해 매핑률이 떨어져 여러 가지 조건으로 대상자를 선별할 때 인원수가 상당히 적게 나올 가능성이 있어 아직은 연구용으로 활용하기 위한 데이터베이스로는 부적합한 상태이다. 또한 5.3.0_02 버전까지 공개되어 있는 매핑코드에 대한 정보가 5.3.1버전부터는 업체로부터 비공개로 되어 있어 추후 CDM 데이터베이스 통한 연구시에 표준 용어에 대한 지식 필수일수 있다.

CDW데이터베이스 및 웹인터페이스인 iDEA에 대해서는 본 연구의 대상자 선별 조건으로 각각 DW와 CDW 데이터베이스로부터 대상자를 산출시 동일한 수가 산출되어 구DW의데이터는 CDW로의 이전이 잘 이루어져 있을 것으로 볼수 있다. 또한 iDEA로 CDW 데이터 베이스로부터 동일한 조건의 대상자를 산출할 때 여러 조건들에 대해서

동일수의 인원이 구해졌다. 이를 미루어볼 때 iDEA 역시 CDW데이터베이스로부터 자료를 추출하고 검색에 있어 좋은 구현성을 보임을 확인할 수 있었다. 하지만, iDEA의 기능적 제한으로 동일 KCD코드 밑으로 세부적인 진단명에 대해서 각각이 그룹화되어, 대상자 그룹의 검색 추출시 특정 개인으로 선택 기능이 없어 모든 경우가 조회되어 특정 개인의 중복이 발생하는 기능적인 제한점이 발견되었다. 또한 대상자의 선별을 위해 여러 개의 조건을 추가할 때 행렬구조의 조건수의 제한으로 중간에 다운로드 및 재업로드과정을 거쳐야 하는 문제가 발견되었다. 더불어 구DW자체가 2011년 전자의무기록(EMR)이 작성되면서 구축된 것으로 EMR 시스템전의 2011년이전의 진료기록과 흡연, 주증상호소 등의 연관된 자료는 활용이 제한적이었다.

당뇨병환자에서 statin제제의 사용에 따른 심혈관질환 발생위험과 사망위험을 보는 실제 분석에서는 CDM데이터 베이스를 활용한 분석결과와 DW를 기반으로 공단자료를 연계한 자료로부터 분석한 결과가 사망위험을 제외하고는 심혈관질환의 위험에서 반대의 결과가 나타났다. 이는 CDM데이터베이스는 일산병원만의 약물사용 및 심혈관질환의 결과만 포함되는 반면에 DW를 기반으로한 공단연계 자료의 경우에 전국단위의 약제와 심혈관질환의 발생의 결과값을 확인함에 있어서 statin 사용군과 비사용군이 서로 배타적인 상황에서 약제사용여부로 인해 그룹이 변경되고 각 군에서의 발생도 달라지므로 위험도에도 영향을 미칠수 있을 것으로 판단된다.

이번 연구는 또한 2020 당뇨병 팩트 시트에서 발표되었던 것처럼 30세이상에서 전체 당뇨병환자의 53.3%에서 LDL-C일 조절되고 있었는데⁴, 이보다도 좀더 높은 statin 비사용군 사용군 에서 각각 53.2% 61%로 사용군에서 LDL-C 100mg/dL미만으로 조절율을 보여 일산병원에서 관리되는 환자들이 좀더 높은 조절율을 보였다.

Statin 비사용군에서 보여준 특징중 하나는 기저치가 LDL-C 105.5mg/dL로 사용군의 134mg/dL에 비해서 확연히 낮았다는 점이다. LDL-C 수치가 많이 높지 않은 군에서 statin제제가 사용되지 않는 경향이 있음을 보여준다. 이는 당뇨병환자를 관리하는 의료진이 다양한 경로를 통해서 LDL-C의 조절의 중요성을 파악하고 있으며, 조절목표인 130mg/dl를 많이 상회하는 경우에는 적극적으로 statin제제를 활용하고 있음을 짐작할 수 있다. 반대로 LDL-C 100mg/dl에 가깝게 유지하는 경우에는 Statin 제제를 사용하는 것을 보유하는 경우도 있음을 알 수 있다.

Statin 사용군에서 심혈관질환 위험의 감소는 경향은 보였으나 통계적으로 유의미하진 않았다. 하지만 statin 사용군에서 비사용군에 비해 전체 사망률이 통계적으로 유의미하

게 감소하였다. 특히 LDL-C 100mg/dL인 군이 비해서 모든 LDL-C 수치의 Statin 사용군의 사망률이 유의미하게 감소하였다. 이는 추후 더 대규모의 연구를 통한 확인이 필요할 수 있지만 100mg/dL 정도로 LDL-C의 수치를 유지하는 경우에도 보다 적극적으로 statin 제제를 당뇨병환자에서 사용하는 것이 모든 원인의 사망을 감소시키는 등의 의미있는 당뇨병환자 관리 전략이 될 수 있음을 보여준다.

하지만, 이번 연구도 몇몇 제한점들은 존재한다. 먼저는 결과변수에 비해서 적은 대상자수를 들 수 있다. 이는 처음 모집단을 선정할 때 세부적인 검사결과를 활용하기 위한 방안으로 일개병원에서 진단받은 당뇨병환자를 대상으로 했기 때문이다. 게다가 최종적인 군을 설정을 위해 statin제제의 연속적인 사용을 조건으로 포함하면서 대상자 모집단으로부터 대상자수가 급감하였다. 또한 심혈관질환 및 사망이라는 비교적 드문 결과값에 비해서 결과값을 보이는 대상자가 매우 적어 분석에 있어 어려움을 겪었다. 이는 추후 CDM 데이터 베이스 및 일개병원과 공단자료를 연계한 자료를 바탕으로 연구를 진행할 때 변수에 대한 노출후 결과값이 오래걸리거나 발생률이 낮은 중대질환에서 결과변수의 확보에 제한이 될 수 있음을 시사한다.

또한 실제 청구데이터상 반영되지 않는 환자관련 요인 및 환자위험에 따른 임상적 판단이 제대로 반영되지 못했다는 제한점이 있다. 실제 환자를 대하는 의료진들이 약물 투여를 결정할 때는 진료지침에서 권장하는 목표치와 현재 환자의 LDL-C 수치는 물론, 환자의 약물순응도, 환자와의 관계, 환자가 가지고 있는 가족력 및 심혈관질환 발생의 위험예측 등 다양한 요인들이 고려된다. 그러나 이러한 부분들을 병원의 데이터베이스나 병원청구코드만으로는 다 알 수 없기 때문에 statin제제 사용에 관련한 분석에 제한점이 있다고 할 수 있다.

마지막으로 실제 임상현장에서의 더 많은 다양한 변수들이 존재한다는 점이다. 실제 중요한 결과들이 도출되는 연구들을 상당부분 무작위대조군연구 (Randomized controlled trial, RCT)에 기인한 경우가 많다. 하지만 변수들이 실험군 대조군에서 잘 통제되는 RCT와 실제 임상현장과는 다른 변수들이 존재한다. 약물의 용량이나 종류가 통제되는 RCT와 달리 환자에게의 반응정도, 부작용, 약물순응도 등에 따라 약물을 다양하게 선택하게 된다. 경우에 따라서는 부작용 등으로 인해서 약물의 종류를 변경하기도 하고 환자의 생활습관 변화 등에 따라서 약물의 용량을 변경하거나 환자의 요구에 따라서 일시적으로 약물을 중단하는 시도를 해보기도 한다. 게다가 병원이동 및 진료과 변경 등의

변수도 가지고 있어 본 연구에서는 이를 반영할수 없었다. 향후 당뇨병환자에서 보다 대규모 환자군을 대상으로한 연구 및 고령의 환자에서 statin 제제 사용 효과 등은 더 연구해볼 필요가 있다.

제2절 제언

1. CDM 데이터베이스 활용을 위한 제언

- 1) 다양한 형태 디자인의 연구시도 및 연구 디자인 적용을 통한 데이터베이스 오류 확인 및 수정

일산병원의 CDM 데이터베이스는 당분간 NIMC5.3.0_02 버전을 연구실행의 주된 데이터베이스로 해야겠지만, 향후에는 5.3.1 weekly 버전으로 진행될 것이다. 현재 02버전, weekly 버전 모두 현재 지속적으로 오류의 발견 및 수정이 이루어지고 있다. 하지만 이러한 과정은 지속적인 연구과제를 통한 활용을 통해서만 가능하다. 특히 어느 한 연구가 CDM데이터의 모든 도메인에 있는 데이터를 활용하는 것이 아니기 때문에 다양한 연구주제를 가지고 보다 많은 연구자들이 CDM 데이터베이스를 활용해보고 그 속에서 지속적으로 오류를 발견 수정해나가는 과정이 필요한 것이다.

- 2) CDM 데이터 베이스의 한계와 효용에 대한 인식필요

CDM 데이터베이스는 분산형 연구망을 통한 여러 병원과의 협력연구가 가능하다는 특징을 가지고 있다. 하지만 근본적으로는 국가단위의 결과값을 알기 어려운 한 병원 단위의 자료이다. 한 병원안에서 발생하는 자세한 진료행위나 검사결과를 활용할수 있는 장점이 있기는 특정개인에게 발생하는 다른 병원에서의 질병발생 등에 대해서는 알기 어렵다는 단점이 존재한다. 특히 2차 종합병원, 3차 상급병원 등으로 구별되고 환자가 원하면 쉽게 병원의 이동이 가능한 우리나라의 의료시스템의 현실에서 이는 매우 중대한 제한점이 될 수 있다. 연구에 있어 변수의 노출에 있어 너무 길어서 중간에 노출 중단 등의 변수가 많을 경우, 노출변수에 비해 결과변수의 발생이 드물거나 결과변수의 발생까지 긴 시간이 필요한 경우, 질환의 특성상 지역내외로 다른 병원으로 전원할 가능성이 많은 경우에는 연구 주제로 적합하지 않을 수 있다.

실제로 최근에 CDM과 분산연구망을 활용하여 COVID-19 백신의 부작용 발생에 대한 연구, 골다공증을 가진 여성에서 alendronate 와 raloxifen의 안전성과 효과분석연구, 안과영역에서 유리체강내 주입술과 그 부작용에 대한 연구 등을 시행하였

다²³⁻²⁵. 이렇게 현재까지의 연구는 주로 약물의 안전성 및 부작용 등 노출로부터 길지 않은 시간내에 결과값이 나오는 연구가 CDM 연구의 주요한 주제가 되는 현실이다. 이를 감안하여 연구자들은 연구주제를 설정하고 계획할 필요가 있다. 반면에 국가 단위의 CDM 데이터 베이스로는 국민건강보험 표준코호트 데이터에 기반한 CDM 데이터베이스도 존재하며 이를 이용한 연구도 진행되어 발표되었던 적도 있다²⁶. 이에 추후에 사용이 가능해 진다면 해당 데이터베이스를 활용한 전국단위의 노출변수를 가진 연구와 이를 바탕으로 한 외국데이터와의 협업도 고려해볼만하다.

3) 향후 CDW을 통한 CDM 데이터베이스의 고도화

일산병원에서는 CDW시스템을 데이터소스로 하여 CDM 데이터베이스가 자동적으로 자료가 탑재되고 CDW의 다양한 비정형 데이터들도 함께 탑재되는 고도화작업을 준비하고 있다. 특히 응급센터의 국가응급진료정보망, 알려지 및 이상반응 정보, 장비인터페이스 결과, 건강검진 문진 자료 등의 비정형 자료 및 기존 자료의 추가 구축을 진행 중이다. 이러한 작업이 잘 이루어진다면 국내에서도 선도적으로 변화하고 국내외의 다른 병원 CDM 데이터베이스가 가지고 있지 않은 CDM에 탑재 가능한 데이터 모형을 구축하는 계기가 될 수 있으리라 생각한다.

또한 여러 병원에서 CDM에서 확장 가능한 여러 데이터모형들을 생산해내고 있으며 이러한 확장모델에 대한 논의들도 계속되고 있다. 따라서 OHDSI Korea 네트워크 및 분산형 바이오헬스 빅데이터 사업단과의 긴밀한 관계속에서 변화에 지체되지 않고 계속해서 변화들에 대해서 유용성을 검토하고 도입여부를 고민해 나가야 할 것이다.

4) CDM 연구를 위한 지원

연구자들이 보다 지속적으로 CDM 데이터베이스에 관심을 가지고 다양한 연구를 시도해보기 위해서는 보다 꾸준한 지원이 필요하다.

특히 5.3.1 weekly 버전부터는 매핑코드가 지원되지 않는 상황에서 OMOP CDM 데이터베이스를 다루며 연구설계를 하고 진행하기 위해서는 표준용어를 통한 연구디자인 및 진행에 대한 교육이 필요하다. 보통의 연구자들이 익숙한 KCD7 또는 ICD-10 코드와는 달리 CDM에서는 이와 연관된 표준용어를 Athena를 통해 탐색하고 연구자가 원하는 진단을 상위구조, 하위구조 속에서 잘 선택할수 있어야 한다. 또한 약제처치에 대한 코드에 대해서도 이에 해당되는 표준용어의 concept ID를 잘 확인할수 있어야 한다. 또한 일산병원 ATLAS외에도 OMOP-CDM 네트워크상에서 많이 사용

하는 open ATLAS, gitHub 등에서 기존 연구자들이 사용했던 결과물들을 잘 활용하는 방법 등에 대한 교육도 필요하다. 또한 단일 기관의 연구가 분산형 연구망을 통해서 다기관연구로 발전하기 위해서는 R을 통한 분석프로세스 구성 및 확산이 필수적이다. 따라서 R프로그램에 대한 실제 활용을 위한 교육 뿐만 아니라 이에 위한 지원도 필요하다고 하겠다.

2. CDW 데이터베이스 활용을 위한 제언

1) 연구자와 CDW 구축단과의 지속적인 협업

현 CDW 시스템이 구축되기까지 CDW 구축단을 중심으로 많은 노력을 기울여져 왔다. 이제는 1,2차 구축사업이 끝나가는 시점에서 앞으로 고도화하는 작업을 앞두고 있다. 이제는 본 연구과정에서 기능적인 오류 등을 개선하고 처음에 계획했던 바와 같이 레지스트리, 비정형데이터 및 외부데이터로의 확장 등의 고도화에 매진할 시기이다. 하지만 단순히 이러한 과정은 시스템자체에서 문제를 찾는 과정이 아닐 것이다. 궁극적으로는 연구자들이 연구를 실행함에 있어 잘 활용할 수 있도록 하는 것이 중요하기 때문에 계속해서 연구자들과의 교류를 중요하다. 특히 연구자들이 실제로 연구과제를 현 CDW와 iDEA를 통해서 적용해보고 그 과정에서 발견되는 오류와 개선점을 피드백을 통해서 전달하고 이를 다시 시스템에 적용하는 선순환적인 구조가 필요하다. 현재 진행중인 자료의 추가적인 구축에 대해서도 응급센터의 국가응급 진료정보망, 알리지 및 이상반응 정보, 장비인터페이스 결과, 건강검진 문진 자료들이 추가될 때 연구자들이 이를 잘 활용할수 있도록 알리고 지원하는 것도 필요할 것이다.

2) 연구자를 위한 연구지원

CDW가 연구자친화적인 구조를 가지고 있고 iDEA를 통해서 구DW보다 보다 손쉽게 대상자를 산출하고 그 특성들을 확인할 수 있다고 해도 연구가 잘 진행되기 위해서는 계속해서 관련부서의 지원이 필수적이다. 특히 iDEA의 기능적인 제한점들이 아직 까지 존재하는 상황에서 iDEA의 사용과 이를 통한 대상자 선별에 있어서 의무기록부의 지원을 필수적이다. 또한 이렇게 정리된 대상자를 바탕으로 연구가 끝까지 진행되기 위해서는 연구소 산하 연구분석부의 지원도 함께 해야할 것이다.

3) CDW 데이터 베이스의 한계와 효용에 대한 인식

CDW 데이터 베이스가 병원에서 이루어지는 상세한 진료내역과 검사결과 등을 가지고 있는 것은 사실이나 그에 반해 일개병원의 데이터라는 한계점 또한 가지고

있다. 일개병원으로서의 연구의 결과가 일반화되기 쉽지 않다는 한계점이 존재한다는 것이다. 특히 일개병원의 자료는 2차 종합병원, 3차 상급병원 등으로 구별되고 환자가 원하면 쉽게 병원의 이동이 가능한 우리나라의 의료시스템의 현실에서 이는 매우 중대한 제한점이 될수 있다. 따라서, 일개병원안에서도 충분한 대상자를 확보할 만한 질병군 또는 이를 바탕으로 레지스트리 연구가 좋은 연구방법이 될수 있다. 특히 최근에 CDW를 바탕으로 한 연구를 살펴보면, CDW 데이터베이스로부터 Triple-negative breast cancer에 대한 CDW 자료로부터 분석한 연구가 있었으며²⁷, CDW 기반 레지스트리로부터 Stage III 비소세포폐암환자에서 10년간 환자추적한 결과를 발표한 연구도 있었다²⁸.

이렇게 연구자들은 공단데이터가 아닌 병원데이터만의 상세한 검사 및 진료정보를 바탕으로 기존에 청구자료만으로는 어려웠던 특정 환자군을 대상으로 한 연구를 고려해볼 만하다. 또한 CDW에서 구축되기로한 레지스트리도 잘 관리 및 활용된다면 좋은 연구를 위한 원천데이터가 될 것이다. 나아가 본 연구에서 진행했던 것처럼 CDW를 바탕으로 한 자료에 일산병원의 강점으로 공단자료가 연계되어 연구에 활용된다면 단일 기관 자료로서의 한계를 좀더 보완할 만한 연구가 진행될 수 있으리라 본다.

참고문헌

참고문헌

1. Grimes DA, Schulz KF. An overview of clinical research: The lay of the land. Lancet [Internet]. 2002 Jan 5;359(9300):57-61.
2. Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE. Medical data mining: Knowledge discovery in a clinical data warehouse. Proc AMIA Annu Fall Symp [Internet]. 1997:101-5.
3. 이신미, 박래웅. Basic Concepts and Principles of Data Mining in Clinical Practice. [Internet]. ; 2009. Available from: <http://repository.ajou.ac.kr/handle/201003/9102>
http://www.e-hir.org/journal/view.html?uid=20&start=&sort=Regnum-0&scale=50&key=&oper=&key_word=&year1=&year2=&Vol=015&Num=02&PG=&book=Journal&mod=vol&sflag=&sub_box=Y&aut_box=Y&sos_box=&pub_box=Y&key_box=&abs_box=&year=
4. Korean Diabetes association. Diabetes Fact sheet in korea 2020. ; 2021.
5. Park IB, Baik SH. Epidemiologic characteristics of diabetes mellitus in korea: Current status of diabetic patients using korean health insurance database. Korean Diabetes J [Internet]. 2009 Oct;33(5):357-62.
6. Turner RC, Millns H, Neil HA, Stratton IM, Manley SE, Matthews DR, Holman RR. Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United kingdom prospective diabetes study (UKPDS: 23). BMJ [Internet]. 1998 Mar 14;316(7134):823-8.
7. Grundy SM, Cleeman JI, Merz CN, Brewer HB,Jr, Clark LT, Hunninghake DB, Pasternak RC, Smith SC,Jr, Stone NJ, National Heart, Lung, and Blood Institute, American College of Cardiology Foundation, American Heart Association. Implications of recent clinical trials for the national cholesterol education program adult treatment panel III guidelines. Circulation [Internet]. 2004 Jul 13;110(2):227-39.
8. 대한당뇨병학회. 2015 당뇨병 진료지침 Treatment Guideline for Diabetes. ; 2015.
9. 보건복지부, 한국보건의료정보원. 2020년 보건의료정보화 실태조사 결과보고서. ; 2021.
10. Inmon WH. The data warehouse and data mining. [Internet]. ; 1996. Available from: <https://link.gale.com/apps/doc/A18993844/AONE?u=anon~c4ec80c1&sid=>

googleScholar&xid=e2812eae

11. Yih WK, Kulldorff M, Fireman BH, Shui IM, Lewis EM, Klein NP, Baggs J, Weintraub ES, Belongia EA, Naleway A, Gee J, Platt R, Lieu TA. Active surveillance for adverse events: The experience of the vaccine safety datalink project. *Pediatrics* [Internet]. 2011 May;127 Suppl 1:54.
12. Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, Bonhoeffer J, Schuemie M, van der Lei J, Sturkenboom M. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: Why and how? *J Intern Med* [Internet]. 2014 Jun;275(6):551-61.
13. Toh S, Reichman ME, Graham DJ, Hampf C, Zhang R, Butler MG, Iyer A, Rucker M, Pimentel M, Hamilton J, Lendle S, Fireman BH, Mini-Sentinel Saxagliptin-AMI Surveillance Writing Group. Prospective postmarketing surveillance of acute myocardial infarction in new users of saxagliptin: A population-based study. *Diabetes Care* [Internet]. 2018 Jan;41(1):39-48.
14. Clair Blacketer. The Common Data Model in *The book of OHDSI: observational health data sciences and informatics*, Observational Health Data Sciences and Informatics, editor. ; 2021.
15. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* [Internet]. 2011 [cited 4/30/2022];19(1):54-60. Available from: <https://doi.org/10.1136/amiajnl-2011-000376>
16. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li Y, Stang PE, Madigan D, Ryan PB. Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers. *Stud Health Technol Inform* [Internet]. 2015;216:574-8. Available from: <https://pubmed.ncbi.nlm.nih.gov/26262116><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4815923/>
17. Rosenquist KJ, Fox CS. Mortality Trends in Type 2 Diabetes In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, Gregg EW, Knowler WC, Barrett-Connor E, Becker DJ, Brancati FL, Boyko EJ, Herman WH, Howard BV, Narayan KMV, Rewers M, Fradkin JE, editors. *Diabetes in America*. [Internet]. 3rd

- ed. Bethesda (MD): ; 2018 DOI: NBK568010 [bookaccession]
18. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med* [Internet]. 1998 Jul 23;339(4):229-34.
 19. Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, Neil HA, Livingstone SJ, Thomason MJ, Mackness MI, Charlton-Menys V, Fuller JH, CARDS investigators. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the collaborative atorvastatin diabetes study (CARDS): Multicentre randomised placebo-controlled trial. *Lancet* [Internet]. 2004;364(9435):685-96.
 20. Nichols GA, Joshua-Gotlib S, Parasuraman S. Independent contribution of A1C, systolic blood pressure, and LDL cholesterol control to risk of cardiovascular disease hospitalizations in type 2 diabetes: An observational cohort study. *J Gen Intern Med* [Internet]. 2013 May;28(5):691-7.
 21. Kim MK, Han K, Joung HN, Baek KH, Song KH, Kwon HS. Cholesterol levels and development of cardiovascular disease in koreans with type 2 diabetes mellitus and without pre-existing cardiovascular disease. *Cardiovasc Diabetol* [Internet]. 2019 Oct 22;18(1):139-9.
 22. 김청수, 이성원, 박래웅. 국내 공통데이터모델 현황 및 품질기준 개발연구. *아주대학교 의료정보학과*; 2021.
 23. Kim Y, Tian Y, Yang J, Huser V, Jin P, Lambert CG, Park H, You SC, Park RW, Rijnbeek PR, Van Zandt M, Reich C, Vashisht R, Wu Y, Duke J, Hripcsak G, Madigan D, Shah NH, Ryan PB, Schuemie MJ, Suchard MA. Comparative safety and effectiveness of alendronate versus raloxifene in women with osteoporosis. *Sci Rep* [Internet]. 2020 Jul 6;10(1):11115-8.
 24. Mun Y, You SC, Lee DY, Kim S, Chung YR, Lee K, Song JH, Park YG, Park YH, Roh YJ, Woo SJ, Park KH, Park RW, Yoo S, Chang DJ, Park SJ. Real-world incidence of endophthalmitis after intravitreal anti-VEGF injections in korea: Findings from the common data model in ophthalmology. *Epidemiol Health* [Internet]. 2021;43:e2021097.
 25. Li X, Ostropolets A, Makadia R, Shaoibi A, Rao G, Sena AG, Martinez-Hernandez

- E, Delmestri A, Verhamme K, Rijnbeek PR, Duarte-Salles T, Suchard M, Ryan P, Hripcsak G, Prieto-Alhambra D. Characterizing the incidence of adverse events of special interest for COVID-19 vaccines across eight countries: A multinational network cohort study. medRxiv [Internet]. 2021 Apr 17
26. Seo SI, Park CH, You SC, Kim JY, Lee KJ, Kim J, Kim Y, Yoo JJ, Seo WW, Lee HS, Shin WG. Association between proton pump inhibitor use and gastric cancer: A population-based cohort study using two different types of nationwide databases in korea. Gut [Internet]. 2021 Nov;70(11):2066-75.
27. Kim H, Kim HJ, Kim H, Kim HR, Jo H, Hong J, Kim R, Kim JY, Ahn JS, Im YH, Lee SK, Kim H, Shin SY, Park YH. Real-world data from a refractory triple-negative breast cancer cohort selected using a clinical data warehouse approach. Cancers (Basel) [Internet]. 2021 Nov 21;13(22):5835. doi: 10.3390/cancers13225835.
28. Jung HA, Sun JM, Lee SH, Ahn JS, Ahn MJ, Park K. Ten-year patient journey of stage III non-small cell lung cancer patients: A single-center, observational, retrospective study in korea (realtime automatically updated data warehouse in health care; UNIVERSE-ROOT study). Lung Cancer [Internet]. 2020 Aug;146:112-9.

부 록



부록

부록표 1. PCI 및 CABG의 CDM내 Concept ID

원내처방코드	EDI코드	영문처방명	concept ID
8495300	O1642	Aorto-Coronary Bypass(2 Site)	2725227
8495000	O1641	Aorto-Coronary Bypass(1 Site)	4106548
8820200	O1641	CABG(LIMA - LAD)	4106548
8817200	O1641	CABG 1 Graft	4106548
8495200	OA641	Off Pump CABG(Simple: 1site)	4305852
8818300	OA641	OPCAB 1 Graft	4305852
8819000	O1640	Coronary Artery Bypass with Autogenous Graft, two Grafts	2725227
8817300	O1640	CABG 2 Grafts	2725227
8495400	OA640	Off Pump CABG(Simple: 2 Site)	4305852
8818400	OA640	OPCAB 2 Graft	4305852
8820500	O1648	3CABG(Y Graft: LIMA-dLAD, LIMA-RA to OM / Aorto-PDA with fein)	2725289
8495310	O1648	Aorto-Coronary Bypass(3 Site)	2725289
8819200	O1648	Coronary Artery Bypass with Autogenous Graft, three Grafts	2725289
8817400	O1648	CABG 3 Grafts	2725289
8821000	O1645	Coronary Artery Bypass with Autogenous Graft of Internal mammary Artery, single Graft	4020466
8496000	OA647	Off Pump CABG(Complex)[관상동맥우회로술 기왕력 있는 경우]	4020216
8817000	OA647	Redo OPCAB	4020216
8821900	OA631	Angioplasty of Coronary Artery	4050713
SD0141	HA681	ANGIOGRAM (GRAFT: FEE ONLY)(2번째혈관~)	4098217
SD0530	M6564	PTCA+Atherectomy+Stent(Additional Vessel)	4283892
SD0593	M6566	P.T.Placement of Intracoronary Stent, PCI of Chronic Total Occlusion	43527998
SD052	M6561	P.T.Placement of Intracoronary Stent(Single Vessel)	4283892
SD059	M6553	PTCA of Culprit lesion in acute myocardial infarction	4006788
SD0592	M6565	P.T.Placement of Intracoronary Stent, PCI of Culprit lesion in acute myocardial infarction	43527998
SD053	M6562	P.T.Placement of Intracoronary Stent(Additional Vessel)	4283892
SD0594	M6567	PTCA+Atherectomy+Stent, PCI of Chronic Total Occlusion	43528004
SD050	M6551	PTCA(Single Vessel)	4006788
SD051	M6552	PTCA(Additional Vessel)	4006788
SD0520	M6563	PTCA+Atherectomy+Stent(Single Vessel)	4283892
SD0591	M6554	PTCA of Chronic Total Occlusion	37017357

부록표 2. Brain CT 및 MRI의 CDM내 Concept ID

원내처방코드	EDI코드	영문처방명	Concept_id
TC441	HA441	LIMITED CT(BRAIN)	4145739
TC451	HA451	COMPUTER TOMOGRAPHY (BRAIN)	4244986
TC461	HA461	COMPUTER TOMOGRAPHY (BRAIN: 조영제사용)	4288538
TC471	HA471	3D BRAIN COMPUTER TOMOGRAPHY	4145739
RM101	HE201006	Brain MRI with CE-촬영료	2789359
RM102	HE102006	HIPPOCAMPUS MRI-촬영료	2789361
RM106F	HF101006	DIFFUSION WEIGHTED IMAGE MRI	4260263
RM108F	HE101006	BRAIN MRI-촬영료	2789361
RM109F	HE101006	BRAIN MRI-촬영료	2789361
RM111F	HE101006	BRAIN MRI-촬영료	2789361
RM132	HE135006	BRAIN MRA(VENOGRAM)-촬영료	4082846
RM150	HF103006	T-M Joint MRI	4082847
RM1501	HF203006	MRI SPECTROSCOPY(기본검사와 동시)	4082847
RM151	HF101006	DIFFUSION WEIGHTED IMAGE MRI	4260263
RM1511	HF201006	DIFFUSION WEIGHTED IMAGE MRI(기본검사와 동시)	4260263
RM152	HF102006	PERFUSION IMAGE MRI	4082989
RM1521	HF202006	PERFUSION IMAGE MRI(기본검사 동시)	4082989
RM154	HF105006	BRAIN DYNAMIC MRI	4013636
RM180	HE401006	BRAIN MRI(LIMITED)	4199388
RM196	HE201006	BRAIN MRI (NAVIGATION: 조영제사용)	2789359
RM230F	HE136006	NECK MRA-촬영료	36713033

공동연구보고서 NHIMC-2021-PR-022

당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM (commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교

발행일	2022년 8월 31일
발행인	강도태, 김성우
편집인	이해중, 이천균
발행처	국민건강보험공단 건강보험연구원, 일산병원 연구소
주소	강원도 원주시 건강로 32(국민건강보험공단) 경기도 고양시 일산동구 일산로 100(국민건강보험 일산병원)
전화	031) 900-6986, 6985
팩스	031) 900-6999
인쇄처	지성프린팅 (02-2278-2490)



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

당뇨병환자 코호트에서 Statin제제 복용유무 및 LDL 콜레스테롤 수치에 따른 심혈관질환 발생위험 비교연구: CDM(commondatamodel) 및 병원(CDW)-공단 자료 연계한 당뇨병환자 코호트 결과 비교