

# <2024.04 - 건강검진코호트DB Study Design>

작성자 : 이은경

## 1. 분석 집단 정의

### 1-1) 모집단 정의

: 2004년 이후, incidence T2DM 진단받은 객체 (*prevalent T2DM* → *incidence T2DM* 으로 변경)

: 2002~2003년 대상, wash-out 진행

Wash-out 조건	2002~2003년 검진DB & 생애전환검진DB	공복혈당이 126 이상 OR 당뇨병 과거력(문진) 존재
	2002~2003년 진료DB - 20T, 60T	상병 변수(주/제 1-부상병)에 “E11~”, “E12~”, “E13~”, “E14~” ICD-10 코드 존재 <b>AND</b> 처방된 약물 중 당뇨병 치료제 처방 이력이 존재

: wash-out에서 제외되지 않은 객체 대상으로 포함 조건 충족 여부(incidence T2DM) 판단

: 2004년 이후 기록 이용

포함 조건	첫 번째 조건	검진DB에 BMI, WAIST, 공복 혈당 기록이 적어도 한 건 이상 존재해야 함. + 진료기록, 처방 기록 적어도 한 건 이상 존재해야 함.	
	두 번째 조건	검진DB & 생애전환검진DB	공복 혈당 수치가 126 이상
		진료DB - 20T, 60T	상병 변수(주/제 1-부상병)에 “E11~”, “E12~”, “E13~”, “E14~” ICD-10 코드 존재 <b>AND</b> 처방된 약물 중 당뇨병 치료제 처방 이력이 존재

→ 첫 번째 조건을 만족하는 개체들 중, **두 번째 조건 중 하나만 만족해도 모집단에 포함.**

(당뇨병 치료제 ATC code)

name	use	code
insulin	insulin	A10A~
경구당뇨약	경구당뇨약	A10B~
Metformin	경구당뇨약	A10BA~
meglitinides	경구당뇨약	A10BX02, A10BX03, A10BX08
SGLT2 inhibitors	경구당뇨약	A10BK~
thiazolidinediones	경구당뇨약	<a href="#">A10BG~</a>
sulfonylureas	경구당뇨약	A10BB~, A10BC~
GLP-1 receptor agonists	경구당뇨약	<a href="#">A10BJ~</a>
DPP-4 inhibitors	경구당뇨약	A10BH~
α-glucosidase inhibitors	경구당뇨약	A10BF~

## 1-2) 포함 조건 만족하는 객체 기준, Disease onset date, Cohort entry date 정의

: Cohort entry date와 Disease onset date 구분해서 파악

Disease Onset Date	Cohort entry date
<u>(후보)</u> - 검진DB 기준으로 공복혈당이 126 이상인 첫 시점 - 진료DB 기준으로 상병 코드에 당뇨병 코드가 기록 & 처방전 table 기준으로 처방 약물이 당뇨 경구 치료제인 첫 시점 → 두 후보 중 <u>가장 이른 시점</u>	Disease onset date 이후, 가장 가까운 검진 시점 <u>(조건)</u> - Cohort entry date 이전에 진료, 처방 기록 적어도 한 건 이상 존재해야 함. - Cohort entry date 이후로 BMI가 한 번 이상 추적되어 있어야 함.

## 1-3) 모집단 제외 기준

- 공통 제외 조건

- I) cohort entry date 전, type I 당뇨("E10~")을 진단 받은 객체
- II) cohort entry date 전, 임신 중 당뇨("O24~")을 진단 받은 객체
- III) cohort entry date 전, insulin 처방 이력이 있는 존재(주성분 4자리만 이용)  
 -- 24.04\_Covariate\_목록\_정리.xlsx 파일 참고
- IV) cohort entry date 이후, 2년 내 검진 기록이 없는 객체

- Outcome이 "Mortality" 일 때

: 공통 제외 조건만 사용

- Outcome이 나머지 "Secondary outcome" 일 때

- : 공통 제외 조건 + outcome 과거력을 가진 객체
- : 나머지 secondary outcome은 covariate로 input

## 2. Covariate 정의 — Master Document 확인 / 조작적 정의가 필요한 변수만 기입

### 1) 검진DB 기반 - 음주습관

2002년 ~ 2007년	: “DRNK_HABIT_RSPS_CD”(음주습관) 변수 이용												
	<table><tr><th>2002~2008</th><th>daily frequency</th></tr><tr><td>1</td><td>0</td></tr><tr><td>2</td><td>0.081967213</td></tr><tr><td>3</td><td>0.214285714</td></tr><tr><td>4</td><td>0.5</td></tr><tr><td>5</td><td>1</td></tr></table>	2002~2008	daily frequency	1	0	2	0.081967213	3	0.214285714	4	0.5	5	1
	2002~2008	daily frequency											
	1	0											
	2	0.081967213											
	3	0.214285714											
4	0.5												
5	1												
: 문진 항목 값을 위 표와 대응되게 값 변경													
: 단위를 “daily”로 모두 통일													
2008년 ~ 2017년	“DRNK_HABIT_RSPS_CD” 변수 이용 - 일주일 단위												
	: 단위를 “daily”로 통일 필요												

	<table border="1"> <thead> <tr> <th>2009~2017</th><th>daily frequency</th></tr> </thead> <tbody> <tr><td>1</td><td>0</td></tr> <tr><td>2</td><td>0.142857143</td></tr> <tr><td>3</td><td>0.285714286</td></tr> <tr><td>4</td><td>0.428571429</td></tr> <tr><td>5</td><td>0.571428571</td></tr> <tr><td>6</td><td>0.714285714</td></tr> <tr><td>7</td><td>0.857142857</td></tr> <tr><td>8</td><td>1</td></tr> </tbody> </table> <p>: 변수 값을 위 표와 대응되게 값 변경 (계산법 : 변수 값 / 7)</p>	2009~2017	daily frequency	1	0	2	0.142857143	3	0.285714286	4	0.428571429	5	0.571428571	6	0.714285714	7	0.857142857	8	1
2009~2017	daily frequency																		
1	0																		
2	0.142857143																		
3	0.285714286																		
4	0.428571429																		
5	0.571428571																		
6	0.714285714																		
7	0.857142857																		
8	1																		
2018년 ~ 2019년	<table border="1"> <thead> <tr> <th>2018~2019</th><th>daily frequency</th></tr> </thead> <tbody> <tr><td>1</td><td>FREQUENCY divided by 7</td></tr> <tr><td>2</td><td>FREQUENCY divide by 30.5</td></tr> <tr><td>3</td><td>FREQUENCY divide by 365</td></tr> <tr><td>4</td><td>0</td></tr> </tbody> </table> <p>: 문진 항목의 두 변수("DRK_PER", "DRK_FRQ") 이용해 단위 "daily"로 통일</p> <ul style="list-style-type: none"> <li>- "DRK_PER"가 1인 경우 "DRK_FRQ" 값을 7로 나눈 값을 사용</li> <li>- "DRK_PER"가 2인 경우 "DRK_FRQ" 값을 30.5로 나눈 값을 사용</li> <li>- "DRK_PER"가 3인 경우 "DRK_FRQ" 값을 365로 나눈 값을 사용</li> </ul>	2018~2019	daily frequency	1	FREQUENCY divided by 7	2	FREQUENCY divide by 30.5	3	FREQUENCY divide by 365	4	0								
2018~2019	daily frequency																		
1	FREQUENCY divided by 7																		
2	FREQUENCY divide by 30.5																		
3	FREQUENCY divide by 365																		
4	0																		

## 2) 검진DB 기반 - 운동빈도

2002년 ~ 2007년	<p>"EXERCI_FREQ_RSPS_CD" 변수 이용 - 일주일 단위</p> <p>: 단위를 "daily"로 통일 필요</p> <p>: 변수 값을 아래 표와 대응되게 값 변경 (계산법 : 변수 값 / 7)</p> <table border="1"> <thead> <tr> <th>2002~2008</th><th>per day</th></tr> </thead> <tbody> <tr><td>1</td><td>0</td></tr> <tr><td>2</td><td>0.21429</td></tr> <tr><td>3</td><td>0.5</td></tr> <tr><td>4</td><td>0.78571</td></tr> <tr><td>5</td><td>1</td></tr> </tbody> </table>	2002~2008	per day	1	0	2	0.21429	3	0.5	4	0.78571	5	1						
2002~2008	per day																		
1	0																		
2	0.21429																		
3	0.5																		
4	0.78571																		
5	1																		
2008년 ~ 2017년	<p>: 운동 관련 위 세 가지 변수 모두 이용</p> <p>→ {(격렬한 운동 변수 값 + 중간정도 운동 변수 값 + 걷기 운동 변수 값) / 3} / 7</p> <table border="1"> <thead> <tr> <th>2009~2017</th><th>per day</th></tr> </thead> <tbody> <tr><td>0</td><td>0</td></tr> <tr><td>1</td><td>0.14286</td></tr> <tr><td>2</td><td>0.28571</td></tr> <tr><td>3</td><td>0.42857</td></tr> <tr><td>4</td><td>0.57143</td></tr> <tr><td>5</td><td>0.71429</td></tr> <tr><td>6</td><td>0.85714</td></tr> <tr><td>7</td><td>1</td></tr> </tbody> </table>	2009~2017	per day	0	0	1	0.14286	2	0.28571	3	0.42857	4	0.57143	5	0.71429	6	0.85714	7	1
2009~2017	per day																		
0	0																		
1	0.14286																		
2	0.28571																		
3	0.42857																		
4	0.57143																		
5	0.71429																		
6	0.85714																		
7	1																		
2018년 ~ 2019년	: "_FRQ"로 끝나는 운동 관련 세 가지 변수 모두 이용																		

	<p>-&gt; {(격렬한 운동 변수 값 + 중간정도 운동 변수 값 + 근력 운동 변수 값) / 3} / 7</p> <table border="1"> <thead> <tr> <th>2018~2019</th><th>per day</th></tr> </thead> <tbody> <tr><td>0</td><td>0</td></tr> <tr><td>1</td><td>0.14286</td></tr> <tr><td>2</td><td>0.28571</td></tr> <tr><td>3</td><td>0.42857</td></tr> <tr><td>4</td><td>0.57143</td></tr> <tr><td>5</td><td>0.71429</td></tr> <tr><td>6</td><td>0.85714</td></tr> <tr><td>7</td><td>1</td></tr> </tbody> </table>	2018~2019	per day	0	0	1	0.14286	2	0.28571	3	0.42857	4	0.57143	5	0.71429	6	0.85714	7	1
2018~2019	per day																		
0	0																		
1	0.14286																		
2	0.28571																		
3	0.42857																		
4	0.57143																		
5	0.71429																		
6	0.85714																		
7	1																		

3) 운동지속 기간 -> 추가로 생성하는 변수

: 운동시간을 나타내는 단위로 “mets”가 있음

2002년 ~ 2007년	문진 항목에 세분화 되어있지 않던 시기이므로 2009년 ~ 2017년 기록에서 가장 과거의 기록 가져와 “운동 지속 기간” 변수 생성																					
2008년 ~ 2017년	<div>2)에서 제시한 운동 관련 변수 중 “격렬한 운동”, “중간정도 운동” 변수만 이용 (걷기 운동 관련 변수는 2018년 ~ 2019년에 측정되지 않으므로)</div> <div><div>- 운동량에 따른 Met 기준</div><table><tr><td>Moderate intensity exercise</td><td>3 mets</td></tr><tr><td>Vigorous activity</td><td>6 mets</td></tr><tr><td>met minutes</td><td>mets*minutes</td></tr></table><div>↓</div><table><tr><td>Weekly met minutes calculation example</td><td>METS * minutes * (weekly days)</td><td></td></tr><tr><td>every day vigorous exercise more than 20min</td><td>6mets*20min*2</td><td>240</td></tr><tr><td>every day moderate exercise more than 30min</td><td>3 mets*30min*3</td><td>270</td></tr><tr><td>every day walk more than 30min</td><td>3mets*30min*4</td><td>360</td></tr><tr><td></td><td>total met minutes per week</td><td>870</td></tr></table><div>: 위의 Met 기준 이용하여 “격렬한 운동”, “중간 정도 운동” 변수 값 이용해 일주일 단위 총 total met minute 계산</div></div>	Moderate intensity exercise	3 mets	Vigorous activity	6 mets	met minutes	mets*minutes	Weekly met minutes calculation example	METS * minutes * (weekly days)		every day vigorous exercise more than 20min	6mets*20min*2	240	every day moderate exercise more than 30min	3 mets*30min*3	270	every day walk more than 30min	3mets*30min*4	360		total met minutes per week	870
Moderate intensity exercise	3 mets																					
Vigorous activity	6 mets																					
met minutes	mets*minutes																					
Weekly met minutes calculation example	METS * minutes * (weekly days)																					
every day vigorous exercise more than 20min	6mets*20min*2	240																				
every day moderate exercise more than 30min	3 mets*30min*3	270																				
every day walk more than 30min	3mets*30min*4	360																				
	total met minutes per week	870																				
2018년 ~ 2019년	<div>2)에서 제시한 운동 관련 변수 중 “격렬한 운동”, “중간정도 운동” 변수만 이용 (근육 운동 관련 변수는 2009년 ~ 2017년에 측정되지 않으므로)</div> <div><div>- 운동량에 따른 Met 기준</div><div>↓</div><table><tr><td>Moderate intensity exercise</td><td>3 mets</td></tr><tr><td>Vigorous activity</td><td>6 mets</td></tr><tr><td>met minutes</td><td>mets*minutes</td></tr></table><div>: 위의 Met 기준 이용하여 “격렬한 운동”, “중간 정도 운동” 변수 값 이용해 빈도 파악</div><div>-&gt; 각 꼬리 문항 (“~HRS”, “_MINS”) 이용해 minutes 파악(이때, “~MINS” 변수 값이 20 이상 이거나 “~HRS”이 결측이 아닌 경우 “격렬한 운동”은 minutes</div></div>	Moderate intensity exercise	3 mets	Vigorous activity	6 mets	met minutes	mets*minutes															
Moderate intensity exercise	3 mets																					
Vigorous activity	6 mets																					
met minutes	mets*minutes																					

<p>을 “30분”으로, “중간 정도의 운동”은 “20분”으로 시간 파악, (이유는 사람들이 작성할 때 “시간”, “분” 고려하지 않고 작성할 수 있어 시간 과대 추정 가능성 있음 &amp; 2008년 ~ 2017년 자료와 기준 통일)</p> <p>→ 일주일 단위 총 total met minute 계산</p>		
PA_VD_FRQ	[문진7-1]평소1주일간숨이많이차게만드는고강도신체활동을머칠하십니까?주당 일 *고강도 신체활동의 예)달리기,에어로빅,빠른속도로 자전거 타기,건설 현장 노동,계단으로 물건 나르기 등	일
PA_VD_HRS	[문진7-2-1]평소하루에숨이많이차게만드는고강도신체활동을몇시간하십니까?하루에 시간	시간
PA_VD_MINS	[문진7-2-2]평소하루에숨이많이차게만드는고강도신체활동을몇시간하십니까?하루에 분	분
PA_MD_FRQ	[문진8-1]평소1주일간,숨이약간차게만드는중강도신체활동을머칠하십니까?주당 일 *7번 응답에 관련된 신체활동은 제외하고 답해주시요. * 중강도 신체활동의 예)빠르게 걷기, 복식테니스,보통 속도로 자전거 타기, 가벼운 물건 나르기, 청소 등	일
PA_MD_HRS	[문진8-2-1]평소하루에숨이약간차게만드는중강도신체활동을몇시간하십니까?하루에 시간	시간
PA_MD_MINS	[문진8-2-2]평소하루에숨이약간차게만드는중강도신체활동을몇시간하십니까?하루에 분	분
PA_MUSL_FRQ	[문진9]최근1주일동안팔굽혀펴기,윗몸일으키기,아령,역기,철봉등근력운동을한날은머칠하십니까?주당 일	일

#### 4) 검진DB 기반 - 흡연습관

##### i) 2002년 ~ 2008년

변수명	데이터 값	Cleaned Data에서의 값	
흡연상태	1	변함 없음.	
	2		
	3		
흡연지속기간(year가 단위)	1	2.5	: 각 범주가 의미하는 연도 범위의 중간값으로 통일
	2	7	
	3	14.5	
	4	24.5	
	5	30	
하루 흡연량(1pack = 20개비 가정)	1	0.25(5/20)	(전제조건) : 흡연상태가 “current smoker”(3)인 사람만 해당 값 계산 / 단위를 pack으로 맞추기 위해 20으로 나눔
	2	$0.73((10+19)/2)/2$	
	3	$1.48((20+39)/20)/2$	
	4	$2(40/20)$	
Pack years (갑년) : 흡연의 누적 damage 계산	새로 생성하는 변수	하루 흡연량 × 흡연지속기간	(전제조건) : 흡연상태가 “current smoker”(3)인 사람만 해당 값 계산

## ii) 2008년 ~ 2017년

[변수값 조정해야 하는 변수]

흡연상태	SMK_STAT_TYPE_RSPS_CD	CHR(1)	1 : 피우지 않는다. 2 : 과거에 피웠으나 지금은 끊었다 3 : 현재도 피운다
(과거)흡연기간	PAST_SMK_TERM_RSPS_CD	NUM(3)	_년
(과거)하루흡연량	PAST_DSQTY_RSPS_CD	NUM(3)	_개피
(현재)흡연기간	CUR_SMK_TERM_RSPS_CD	NUM(3)	_년
(현재)하루흡연량	CUR_DSQTY_RSPS_CD	NUM(3)	_개피

- For current smoker

(현재) 흡연기간	“years” 기준 — 변동 필요없음
(현재) 하루 흡연량	“#of cigarettes / 20”으로 변환
Pack years(새로 추가)	(현재) 하루 흡연량 × (현재) 흡연기간

- For past smoker

(과거) 흡연기간	“years” 기준 — 변동 필요없음
(과거) 하루 흡연량	“#of cigarettes / 20”으로 변환
Pack years(새로 추가)	(과거) 하루 흡연량 × (과거) 흡연기간

## iii) 2018년 ~ 2019년

SMK_STAT_TYPE_RSPS_CD	흡연상태	1:피우지않는다. 2:과거에피웠으나지금은끊었다 3:현재도피운다
PAST_SMK_TERM_RSPS_CD	(과거)흡연기간	
PAST_DSQTY_RSPS_CD	(과거)하루흡연량	
CUR_SMK_TERM_RSPS_CD	(현재)흡연기간	
CUR_DSQTY_RSPS_CD	(현재)하루흡연량	

[변수값 조정해야 하는 변수]

- For current smoker

(현재) 흡연기간	“years” 기준 — 변동 필요없음
(현재) 하루 흡연량	“#of cigarettes / 20”으로 변환
Pack years(새로 추가)	(현재) 하루 흡연량 × (현재) 흡연기간

- For past smoker

(과거) 흡연기간	“years” 기준 — 변동 필요없음
(과거) 하루 흡연량	“#of cigarettes / 20”으로 변환
Pack years(새로 추가)	(과거) 하루 흡연량 × (과거) 흡연기간

5) 구강검진DB 기반 - Soda 섭취 / 간식 섭취 습관

: 결측률 자체가 높은 Covariate 이므로(이전 파악 결과 참조) 현 시점 분석에서는 사용하지 않는 것으로 결론 지음.

6) 검진DB 기반 - 가족력

: 한 번이라도 가족력이 있다면, **always positive**로 간주(모든 time에 가족력이 존재하는 것으로 판단)  
- 가족력은 유전적인 의미를 가지고 가기 때문.

Heart disease	검진DB의 문진 항목 이용
Diabetes mellitus	검진DB의 문진 항목 이용
Cancer	2002년 ~ 2007년 검진DB 문진 항목에 가족력이 있다고 기록되어 있다면, <b>always positive</b> 로 간주

7) 개인 과거 병력

: 개인 과거 병력은 검진DB, 진료DB 모두에서 확인할 수 있다. (**문진 및 ICD-10 code 모두 이용해서 정의**)

■ ICD-10 code는 24.04\_Covariate\_목록\_정리.xlsx 파일 참고

	Hypertension (고혈압)	Stroke or TIA (뇌졸중)	Myocardial infraction (심근경색 - 혈관 관련)	Heart failure (심장마비 - 근육 관련)	Cancer (암)
검진DB	사용	사용	사용하지 않음	사용하지 않음	사용하지 않음
진료DB	사용	사용	사용	사용	사용

8) Comorbidities

Comorbidities (24.04 Updates)		
diagnosis	diagnosis	code
Asthma	Asthma	J45~, J46~
Chronic obstructive pulmonary disease (COPD)	Chronic obstructive pulmonary disease	J40~, J41~, J42~, J43~, J44~, J47~
Atrial fibrillation	Atrial fibrillation	I48~
Thromboembolism	Thromboembolism	I26~, I27~, I28.0, I28.1, I28.2, I28.3, I28.8, I28.9, I29~
Chronic liver disease	Chronic liver disease	K70~K77~
Chronic Kidney disease	Chronic Kidney disease	N18~
Coronary artery disease (동맥경화, 심장혈관질환)	Coronary artery disease	I20~, I21~, I22~, I23~, I24~, I25~
Peripheral vascular disease	Peripheral vascular disease	I70~, I71~, I72~, I73~, I74~, I77~, I79~
Dementia	Dementia	F00~, F01~, F03~, G30~, G31.00, G31.82
Hyperlipidemia	Hyperlipidemia	E78~
Pneumonia	Pneumonia	J12~, J13~, J14~, J15~, J16~, J17~, J18~

## 9) Medications

: 24.04\_Covariate\_목록\_정리.xlsx 파일 참고

[24.04 Medication list Update]

- "name" 기준 변수 생성

Category	name	ATC code
oral diabetes medication	oral diabetes	A10B~
	Metformin	A10BA~
	sulfonylureas	A10BB~, A10BC~
	others oral diabetes medication	A10BX02, A10BX03, A10BX08, A10BK~, A10BG~, A10BJ~, A10BH~, A10BF~
Non-statin antihyperlipidemic	Non-statin antihyperlipidemic	C10(everythin starting with c10) except (c10AA~, c10BA~, c10BX~, A10BH51, A10BH52)
statin	statin	c10AA~, c10BA~, c10BX~, A10BH51, A10BH52
혈압약 (renin angiotensin system (교감신경) pathway 관련)	Angiotensin converting enzyme inhibitors	C09A~, C09B~
	Angiotensin receptor II blockers	C09C~, C09D~
혈압약 (RAS 제외)	β-blockers	C07~
	Calcium channel blockers	C08~
	Diuretics	C03~
	others hypertension	C01D~, C02D~, C04~, C02A~, C02B~, C02C~
CNS (central nerve system)	Anxiolytics	N05B~
	Antidepressants	N06A~
	Other CNS	N06D~, N05A~
진통제(non-opioid analgesic)	NSAIDs	M01A~
Antithrombotics	Antithrombotics	B01AA~, B01AB~, B01AE~, B01AF~, B01AX~, A01AD05, B01AC~, C07FX02-04, C10BX~, M01BA03, N02AJ02, 07,18, N02BA01,51,71

## 3. Outcome 정의

: Outcome에 따른 population 제외 조건 상이하니 주의 필요

Primary Outcome	all cause death	Outcome 정의 시에는 "ICD-10" code만 활용
Secondary Outcome	Cancer	
	Myocardial infraction	
	Heart failure	
	TIA or Stroke	

## 4. Exposure 정의

: cohort entry date 시점에서의 BMI 값

(Update)

BMI 기준	WHO/WPRO	일본 비만 학회	일본 인간도크학회	한국 (2000년)	한국 (2018년)
<18.5	저체중	저체중	저체중	저체중	저체중
18.5 - 22.9	정상	보통체중	정상	정상(22.9)	정상
23-24.9	정상	보통체중	정상	과체중	비만 전 단계
25.0 - 29.9	과체중	비만 1도	정상 (남 27.5/여 26.1)	비만	1단계 비만
30.0 - 34.9	1 단계 비만	비만 2도		고도비만	2단계 비만
35 - 39.9	2단계 비만	비만 3도			3단계 비만
>40	고도비만	비만 4도			