

< 건강검진코호트DB Analysis Master Document>

[연구가설]

: 당뇨병 진단 받은 이후(Cohort entry date) 체중 감량(Exposure)과 건강 상태(Outcome)간의 인과성이 존재한다.

1. 건강검진코호트DB 파일목록

구분			한글명	파일명
자격유			자격	NHIS_HealS_JK
진료DB	의과_보건기관(T1)	명세서(20t)	명세서_20t_T1	NHIS_HealS_GY20_T1
		진료내역(30t)	진료내역_30t_T1	NHIS_HealS_GY30_T1
		상병내역(40t)	상병내역_40t_T1	NHIS_HealS_GY40_T1
		처방전교부상세내역(60t)	처방전교부상세내역_60t_T1	NHIS_HealS_GY60_T1
	치과_한방(T2)	명세서(20t)	명세서_20t_T2	NHIS_HealS_GY20_T2
		진료내역(30t)	진료내역_30t_T2	NHIS_HealS_GY30_T2
		상병내역(40t)	상병내역_40t_T2	NHIS_HealS_GY40_T2
		처방전교부상세내역(60t)	처방전교부상세내역_60t_T2	NHIS_HealS_GY60_T2
	약국(T3)	명세서(20t)	명세서_20t_T3	NHIS_HealS_GY20_T3
		진료내역(30t)	진료내역_30t_T3	NHIS_HealS_GY30_T3
건강검진DB	일반건강검진(1차)		일반건강검진	NHIS_HealS_GJ
	생애전환기건강진단(1차) (07~15년)		생애전환기건강진단	NHIS_HealS_LJ
	구강검진(일반)		구강검진(일반)	NHIS_HealS_GO
	구강검진(생애) (07~15년)		구강검진(생애)	NHIS_HealS_LO
요양기관DB	요양기관 현황		요양기관	NHIS_HealS_YK

Problem 1) 생애 전환기 건강진단은 2007년부터 시행된 검진이며, 검진 대상자가 만 40세, 60세인 사람들 뿐이다. 이때, 생애 전환기 건강진단 기록도 분석 데이터로 포함시켜야 하는가?

2. Study Population 정의

: 2002년 ~ 2016년 사이 ICD-10 코드 기준으로 당뇨병 진단을 받거나 건강검진 당시 당뇨병 이력이 있는 것으로 파악된 사람 (제일 빠른 진단 날짜가 cohort entry date)

포함 기준	당뇨병 진단	<ul style="list-style-type: none"> - 해당 검사기록은 2002년 ~ 2019년까지 모두 존재. - 공복 혈당 검사 기록 통해 질환 의심 여부 확인 / 진료 table의 상병 기록 확인
	당뇨병 이력	<ul style="list-style-type: none"> - 건강검진DB 문진 항목 통해 확인 i) 2002년 ~ 2008년 : “HCHK_PMH_CD1”, “HCHK_PMH_CD2”, “HCHK_PMH_CD3” (과거병력 변수 - 3개까지 기입 가능하였음) 변수 통해 확인 가능 - “7”이 당뇨병 과거병력을 의미 ii) 2009년 ~ 2017년: “HCHK_HPRTS_PMH_YN” (당뇨병 과거 병력 유무) 변수 통해 확인 가능 - “1”이 해당을 의미 iii) 2018년 ~2019년 : “HCHK_DIABML_PMH_YN” (당뇨병 과거 병력 유무) 변수 통해 확인 가능 - “2”(있음)이 해당을 의미(범주가 1~3) / “1”이 2009년 이후의 병력을 의미(범주가 0~1)
	자격(급여 / 건강보험)	<ul style="list-style-type: none"> - 의료수급권자와 건강보험자격자로 나뉨. - 의료수급권자의 경우 2006년 이전까지의 진료기록이 없음

		- 의료수급권자 의 경우 2012년 이전까지의 건강검진 기록이 없음.
제외 기준	연령	- Proposal 내용 : cohort entry date 당시 연령이 40세 미만인 개체는 40세 이후부터의 기록을 가져옴. - 건강검진 표본 코호트DB 선정 기준 : 2002년 12월 말 당시 만 40세 ~ 79세인 사람 (출처 : 국민건강보험공단 건강검진DB 매뉴얼)

Problem 2) 의료수급권자의 경우 censored된 자료만 얻을 수 있으므로(2006년 이전까지의 진료기록 없음 & 2012년 이전까지의 건강검진 기록 없음), 연구대상자에서 제외함을 고려할 필요가 있다. 그런데, 의료수급권자를 모집단에서 제외한다면 총 모집단의 수가 많이 줄어들 수 있다. 의료수급권자의 기록을 어떻게 다루어야 하는가?
 -- 하단 그림 참조. (하단 그림 출처 : 국민건강보험공단 건강검진DB 매뉴얼)

※ 연도별 자격변동에 따른 의료급여수급권자 수검내역

구분	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
(자격기준) 의료급여수급권자	515	1,785	3,662	6,123	7,482	8,529	8,764	8,061	8,415	8,550	8,324	8,341	8,345	9,488	9,651	10,216	9,948	10,310
일반건강검진 수검자	387	685	349	493	419	435	275	350	240	338	961	1,018	881	1,346	1,116	1,407	2,371	2,887

Note 1) Proposal 작성 당시, cohort entry date 당시 40세 이하인 개체들의 경우는 40세 이후부터의 기록만 가져온다고 정함 --- **연령에 따른 진료기록 포함 여부 고려할 필요가 사라짐.**

Note 2) covariate 변수는 cohort entry date 이전의 모든 기록 확인 통해서 생성됨. -- 만약, cohort entry date가 2002년이면 이전 기록은 확인이 불가능하므로 2002년 기록을 전부 이용한다.

*** 주의사항 :** “당뇨병 이력”의 경우 연도에 따라 확인하는 기준이 다름. --- 잘 살펴봐야 함.

3. Outcome 정의

Primary Endpoint	all-cause death	자격DB - 사망일자 변수(DTH_MDY) 통해 사망 여부 확인 가능
Secondary Endpoint	Risk of MI	진료 레이아웃 20t, 40t에서 상병 변수(MAIN_SICK, SUB_SICK, SICK_SYM)통해 확인
	Heart failure	
	TIA or Stroke	

Note 3) 자격DB에 사망 사유(DTH_CODE1, DTH_CODE2)도 존재

4. Exposure 정의

BMI(체질량 지수)	2002년부터 2019년까지 모두 검진 검사 항목에 포함되어 있음 -- trajectory 파악 가능
WAIST(허리둘레)	2008년에 추가된 항목 으로, 2002년부터 2007년까지의 기록은 파악 불가.

Problem 3) WAIST(허리둘레) 변수의 경우 2008년에 추가된 검사 항목이기 때문에 2002년부터 2007년까지의 trajectory는 파악할 수 없다. 그렇다면, 2002년부터 2007년까지의 WAIST 변수 값을 어떻게 처리해야 하는가?

5. Covariate 정의

i) 인구학적 정보

- 관측 timing : cohort entry date에서 가장 가까운 검진 년도의 기록 이용

Age	자격DB - “AGE” 변수 통해 확인 가능
Gender	자격DB - “SEX” 변수 통해 확인 가능
medical insurance / aid	자격DB - “IPSN_TYPE_CD” 변수 통해 확인 가능 (6부터 의료수급권자)
systolic blood pressure	검진DB - “BP_HIGH” 변수 통해 확인 가능
diastolic blood pressure	검진DB - “BP_LWST” 변수 통해 확인 가능

Question 1) 만약, 의료수급권자를 모집단에서 제외한다면 medical insurance / aid는 covariate로 고려할 필요가 없는 부분인가.

ii) Lifestyle factor

- 관측 timing : cohort entry date에서 가장 가까운 검진 년도의 기록 이용

<p>Ethanol drinking (음주습관)</p>	<p>검진DB 문진 항목 통해 확인 가능</p> <p>i) 2002년 ~ 2008년 : 객관식 형태 (“DRNK_HABIT_RSPS_CD”)</p> <table border="1" data-bbox="710 943 1150 1108"> <tr> <td>CHR(1)</td><td> 1 : (거의)마시지 않는다 2 : 월2~3회정도 마신다 3 : 일주일에 1~2회 마신다 4 : 일주일에 3~4회 마신다 5 : 거의 매일 마신다 </td></tr> </table> <p>ii) 2009년 ~ 2017년 : 일주일 기준 (“DRNK_HABIT_RSPS_CD”)</p> <table border="1" data-bbox="710 1198 1150 1386"> <tr> <td>CHR(1)</td><td> 1 : 0일 2 : 1일 3 : 2일 4 : 3일 5 : 4일 6 : 5일 7 : 6일 8 : 7일 </td></tr> </table> <p>iii) 2018년 ~ 2019년 : “DRK_FRQ” 변수 통해 확인 가능 / 지난 1년 기준</p> <table border="1" data-bbox="710 1509 1163 1637"> <tr> <td></td><td> 1:일주일에 2:한 달에 3:1년에 4:술을 마시지 않는다 </td></tr> </table>	CHR(1)	1 : (거의)마시지 않는다 2 : 월2~3회정도 마신다 3 : 일주일에 1~2회 마신다 4 : 일주일에 3~4회 마신다 5 : 거의 매일 마신다	CHR(1)	1 : 0일 2 : 1일 3 : 2일 4 : 3일 5 : 4일 6 : 5일 7 : 6일 8 : 7일		1:일주일에 2:한 달에 3:1년에 4:술을 마시지 않는다
CHR(1)	1 : (거의)마시지 않는다 2 : 월2~3회정도 마신다 3 : 일주일에 1~2회 마신다 4 : 일주일에 3~4회 마신다 5 : 거의 매일 마신다						
CHR(1)	1 : 0일 2 : 1일 3 : 2일 4 : 3일 5 : 4일 6 : 5일 7 : 6일 8 : 7일						
	1:일주일에 2:한 달에 3:1년에 4:술을 마시지 않는다						
<p>Exercise (운동빈도)</p>	<p>검진DB 문진 항목 통해 확인 가능</p> <p>i) 2002년 ~ 2008년 : 1주일 기준 (“EXERCI_FREQ_RSPS_CD”)</p> <table border="1" data-bbox="710 1818 1163 1991"> <tr> <td>CHR(1)</td><td> 1 : 안한다 2 : 1~2회 3 : 3~4회 4 : 5~6회 5 : 거의 매일 </td></tr> </table>	CHR(1)	1 : 안한다 2 : 1~2회 3 : 3~4회 4 : 5~6회 5 : 거의 매일				
CHR(1)	1 : 안한다 2 : 1~2회 3 : 3~4회 4 : 5~6회 5 : 거의 매일						

ii) 2009년 ~ 2017년 : 1주일 기준 - 분류 세분화

1주_20분이상 격렬한 운동	MOV20_WEK_FREQ_ID	CHR(1)	1 : 0일 2 : 1일 3 : 2일
1주_30분이상 중간정도 운동	MOV30_WEK_FREQ_ID	CHR(1)	4 : 3일 5 : 4일 6 : 5일
1주_총30분이상 걷기 운동	WLK30_WEK_FREQ_ID	CHR(1)	7 : 6일 8 : 7일

iii) 2018년 ~ 2019년 : 분류가 더 세분화 (주관식)

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 답해주십시오. * 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

검진DB 문진 항목 통해 확인 가능

i) 2002년 ~ 2008년 : 객관식

흡연상태	SMK_STAT_TYPE_RSPS_CD	CHR(1)	1 : 피우지 않는다 2 : 과거에 피웠으나 지금은 끊었다 3 : 현재도 피운다
(과거,현재)흡연기간	SMK_TERM_RSPS_CD	CHR(1)	1 : 5년 미만 2 : 5~9년 3 : 10~19년 4 : 20~29년 5 : 30년 이상
(현재)하루흡연량	DSQTY_RSPS_CD	CHR(1)	1 : 반갑미만 2 : 반갑이상~한갑미만 3 : 한갑이상~두갑미만 4 : 두갑이상

ii) 2009년 ~ 2017년 : 분류 세분화

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

iii) 2018년 ~ 2019년 : 분류가 더 세분화 (주관식)

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

Smoking
(흡연습관)

Frequency of eating fruit or vegetables

알 수 있는 변수가 없음.

Frequency of eating soda

구강검진DB 문진 항목 통해 파악 가능 (“OQ_SODA_FRQ” 변수 통해 파악 가능)

	<p>- 2014년부터 추가된 검진 항목 : 하루 기준</p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> 1:먹지않음 2:1번 3:2~3번 4:4번이상 5:모르겠음 </div>
Frequency of eating snack	<p>구강검진DB 문진 항목 통해 파악 가능 ("OQ_Q06" 변수 통해 파악 가능)</p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> 1 : 그렇다 2 : 아니다 3 : 보통이다 </div> <p>- 해당 변수는 2002년 ~ 2008년까지만 문진 항목에 포함되어 있다.</p>

Problem 4) 문진 항목이 2008년, 2018년을 기준으로 많이 변경되어 covariate 변수 생성이 어려움(흡연습관 / 음주습관 / 운동빈도) -- 기준 설립 통해 연도가 달라도 같은 의미를 가지는 covariate 변수 생성이 필요함.

Problem 5) 과일, 채소 섭취 습관의 경우 알 수 있는 변수가 어디에도 없음. / 탄산음료 섭취 빈도, 간식 섭취 빈도의 경우는 일정 시점에서만 파악할 수 있는 변수이다. -- 파악할 수 없는 연도에서는 변수값 어떻게 지정?

iii) Family history

- 관측 timing : cohort entry date에서 가장 가까운 검진 년도의 기록 이용

Heart disease	검진DB 문진 항목 통해 확인 가능 i) 2002년 ~ 2008년 : 객관식 - “2”가 있음을 의미		
	(가족력)심장병유무	FMLY_HDISE_PATIEN_YN	CHR(1)
	(가족력)당뇨병유무	FMLY_DIABML_PATIEN_YN	CHR(1)
	(가족력)암유무	FMLY_CANCER_PATIEN_YN	CHR(1)
Cancer	ii) 2009년 ~ 2017년 : 객관식 - “1”이 해당을 의미		
	(가족력)심장병환자유무	FMLY_HDISE_PATIEN_YN	
	(가족력)당뇨병환자유무	FMLY_DIABML_PATIEN_YN	
	(가족력)기타(암포함)환자유무	FMLY_CANCER_PATIEN_YN	
Diabetes mellitus	iii) 2018년 ~ 2019년 : 객관식 - “2”가 있음을 의미(범주가 1~3) / “1”이 2009년 이후의 가족 병력을 의미 (범주가 0 , 1)		
	FMLY_HDISE_PATIEN_YN	(가족력)심장병환자유무	1:없음 2:있음 3:모름
	FMLY_DIABML_PATIEN_YN	(가족력)당뇨병환자유무	(2009년이후) 0:미해당 1:해당
	FMLY_CANCER_PATIEN_YN	(가족력)기타(암포함)환자유무	

Problem 6) 가족 병력의 경우, 2002년부터 2019년까지 모든 기록을 알 수 있으나 2009년부터 뇌졸중 -> 뇌졸중(중풍), 심장병 -> 심장병(심근경색 / 협심증), 암 -> 기타(암 포함)으로 문진 항목이 변경되었다. 그렇다면, 2009년부터 암 가족력은 식별 가능성이 매우 적다. -- 기타 항목을 어떻게 처리?

iv) Comorbidities

- 관측 timing : cohort entry date 이전 기록 전부에서 병력 확인

Stroke or TIA	<p>방법 1) 진료DB에서 (20t, 40t) 상병 변수 통해 확인("MAIN_SICK", "SUB_SICK", "SICK_SYM" 변수)</p> <p>방법 2) 검진DB의 문진 항목 통해 병력 확인</p> <p>i) 2002년 ~ 2008년 : 객관식("4"가 고혈압, "5"가 심장병, "6"이 뇌졸중, "7"이 당뇨병, "8"이 암을 의미)</p>										
Myocardial infraction	<table border="1"> <tr> <td>(본인)과거병력코드1</td><td>HCHK_PMH_CD1</td></tr> <tr> <td>(본인)과거병력코드2</td><td>HCHK_PMH_CD2</td></tr> <tr> <td>(본인)과거병력코드3</td><td>HCHK_PMH_CD3</td></tr> </table>	(본인)과거병력코드1	HCHK_PMH_CD1	(본인)과거병력코드2	HCHK_PMH_CD2	(본인)과거병력코드3	HCHK_PMH_CD3				
(본인)과거병력코드1	HCHK_PMH_CD1										
(본인)과거병력코드2	HCHK_PMH_CD2										
(본인)과거병력코드3	HCHK_PMH_CD3										
Hypertension	<p>ii) 2009년 ~ 2017년 : 객관식, "1"이 해당을 의미</p> <table border="1"> <tr> <td>(본인)뇌졸중과거병력유무</td><td>HCHK_APOP_PMH_YN</td></tr> <tr> <td>(본인)심장병과거병력유무</td><td>HCHK_HDISE_PMH_YN</td></tr> <tr> <td>(본인)고혈압과거병력유무</td><td>HCHK_HPRTS_PMH_YN</td></tr> </table>	(본인)뇌졸중과거병력유무	HCHK_APOP_PMH_YN	(본인)심장병과거병력유무	HCHK_HDISE_PMH_YN	(본인)고혈압과거병력유무	HCHK_HPRTS_PMH_YN				
(본인)뇌졸중과거병력유무	HCHK_APOP_PMH_YN										
(본인)심장병과거병력유무	HCHK_HDISE_PMH_YN										
(본인)고혈압과거병력유무	HCHK_HPRTS_PMH_YN										
Cancer	<table border="1"> <tr> <td>(본인)당뇨병과거병력유무</td><td>HCHK_DIABML_PMH_YN</td></tr> <tr> <td>(본인)기타(암포함)질환 과거병력유무</td><td>HCHK_ETCDSE_PMH_YN</td></tr> </table>	(본인)당뇨병과거병력유무	HCHK_DIABML_PMH_YN	(본인)기타(암포함)질환 과거병력유무	HCHK_ETCDSE_PMH_YN						
(본인)당뇨병과거병력유무	HCHK_DIABML_PMH_YN										
(본인)기타(암포함)질환 과거병력유무	HCHK_ETCDSE_PMH_YN										
Heart failure	<p>iii) 2018년 ~ 2019년 : 객관식 - "2"가 있음을 의미(범주가 1~3) / "1"이 2009년 이후의 가족 병력을 의미 (범주가 0, 1)</p> <table border="1"> <tr> <td>HCHK_ETCDSE_PMH_YN</td><td>(본인)기타(암포함)질환 과거병력유무</td></tr> <tr> <td>HCHK_APOP_PMH_YN</td><td>(본인)뇌졸중과거병력유무</td></tr> <tr> <td>HCHK_HDISE_PMH_YN</td><td>(본인)심장병과거병력유무</td></tr> <tr> <td>HCHK_HPRTS_PMH_YN</td><td>(본인)고혈압과거병력유무</td></tr> <tr> <td>HCHK_DIABML_PMH_YN</td><td>(본인)당뇨병과거병력유무</td></tr> </table> <div style="margin-left: 20px;"> <p>1:없음 2:있음 3:모름</p> <p>(2009년이후) 0:미해당 1:해당</p> </div>	HCHK_ETCDSE_PMH_YN	(본인)기타(암포함)질환 과거병력유무	HCHK_APOP_PMH_YN	(본인)뇌졸중과거병력유무	HCHK_HDISE_PMH_YN	(본인)심장병과거병력유무	HCHK_HPRTS_PMH_YN	(본인)고혈압과거병력유무	HCHK_DIABML_PMH_YN	(본인)당뇨병과거병력유무
HCHK_ETCDSE_PMH_YN	(본인)기타(암포함)질환 과거병력유무										
HCHK_APOP_PMH_YN	(본인)뇌졸중과거병력유무										
HCHK_HDISE_PMH_YN	(본인)심장병과거병력유무										
HCHK_HPRTS_PMH_YN	(본인)고혈압과거병력유무										
HCHK_DIABML_PMH_YN	(본인)당뇨병과거병력유무										

Problem 7) 병력의 경우, 2002년부터 2019년까지 모든 기록을 알 수 있으나 2009년부터 '뇌졸중 → 뇌졸중(중풍)', '심장병 → 심장병(심근경색 / 협심증)', '암 → 기타(암 포함)'으로 문진 항목이 변경되었다.

그렇다면, 2009년 이전의 심근경색 이력을 어떻게 확인? / 2009년부터 심근경색 & Heart failure, 암 병력을 어떻게 구분? --- 식별 기준 필요

---- 다음 페이지로 ----

v) Medication

- 관측 timing : cohort entry date 이전 기록 전부에서 병력 확인

Non-statin antihyperlipidemic	<div>- 진료DB에서 (30t, 60t) 일반명코드(“GNL_NM_CD”)변수 통해 약물 처방 이력 확인 - “GNL_NM_CD”은 30t, 60t의 공통변수이며 약품의 일반(성분)코드 담고 있다.</div> <div>- 구성은 일련번호(4) + 주성분별 함량일련번호(2) + 투여경로(1) + 제형(2)이다.</div>
Statin	
Angiotension converting enzyme inhibitors	
Angiotension receptor II blockers	
β-blockers	
Calcium channel blockers	
Diuretics	
vasodilators	
Anti-adrenergics	

※ DB Joint key

- 자격DB↔진료DB(20t)↔건강검진DB : 개인일련번호(PERSON_ID)로 연결
- 진료DB의 세부DB(20t, 30t, 40t, 60t) : 청구일련번호(KEY_SEQ)로 연결