
의료정보안내서

목 차

1. 건강보험 청구자료의 이해	1
1.1 건강보험제도	1
1.2 진료비 청구 및 심사	4
1.3 보건의료빅데이터 개방시스템	9
1.4 건강보험 청구자료의 구성	18
1.5 자료이용 시 주의사항	32
1.6 자료신청 시 주의사항	36
1.7 청구자료를 이용한 연구	38
2. 명세서 기반의 분석적 접근	40
2.1 명세서 기반의 질병, 의약품 사용 변수 생성	40
2.2 질병통계 작성	56
2.3 의약품 사용 분석	61
2.4 의료경영지원 관련 지표 생성	70
3. 환자 기반의 분석적 접근	76
3.1 대푯값, 분율, 비율	76
3.2 발생률, 유병률, 사망률	77
3.3 표준화법	81
3.4 상대위험도(relative risk), 오즈비(odds ratio)	84

4. 연구설계	86
4.1 코호트 연구	88
4.2 환자-대조군 연구	90
5. 청구자료를 이용한 성과연구	93
5.1 연구설계	94
5.2 연구대상자 선정	96
5.3 변수 정의	96
5.4 통계분석	97
5.5 결과 해석 시 고려할 점	98
5.6 이차자료원을 이용한 연구 체크리스트	99
6. 사례	101
6.1 개요	101
6.2 자료원	101
6.3 연구설계	102
6.4 연구결과	108
6.5 프로그램 예시	112
참고문헌	120
부록1. RECORD 체크리스트	121
부록2. 명세서 현황 분석	125

1. 건강보험 청구자료의 이해

1.1. 건강보험제도¹⁾

한국의 건강보험은 국가에서 관장하는 전 국민이 통합된 국가 단일보험이며, 의료공급자의 건강보험에 대한 진료나 보험 적용을 의무화하고 있다. 또한 민간이 투자해서 공공의 통제를 받는 민간중심 공급 구조를 갖고 있다.

□ 건강보험 대상자

건강보험 적용 대상자는 국내에 거주하는 국민으로 가입자 및 피부양자를 포함하며, 단 의료급여법에 따라 의료급여를 받는 사람 및 독립유공자예우에 관한 법률 및 국가유공자 등 예우 및 지원에 관한 법률에 따라 의료보호를 받는 사람들은 제외한다. 단 유공자 중 건강보험의 적용을 희망하는 경우 건강보험 적용대상자가 될 수 있다.

□ 건강보험제도

한국의 건강보험제도에는 3개의 조직이 있다. 보건복지부는 건강보험 관련법을 제정하고 제도운동을 관리감독하며, 건강보험제도 운영은 국민건강보험공단과 건강보험심사평가원(이하 심사평가원)이 정부로부터 위탁 받아 운영하고 있다. 2000년 국민건강보험법 시행으로 의료보험의 관리운영체계를 통합하게 되었고 같은 해 국민건강보험공단과 건강보험심사평가원이 출범하게 되었으며, 국민건강보험공단은 보험가입자 자격관리, 보험료 징수, 진료비 지급의 업무를 담당하고 있다.

건강보험심사평가원은 국민건강보험법 제62조에 의해 요양급여비용을 심사하고 요양급여의 적정성을 평가하기 위한 기관으로, 진료비 심사, 의료 질 평가, 진료기준 설정·관리 업무를 수행하고 있다. 요양급여비용을 청구하려는 요양기관은 심사평가원에 요양급여비용의 심사청구를 하며, 심사평가원은 요양급여의 적정성을 평가하여 공단에 통보하고, 심사 내용을 통보받은 공단은 그 내용에 따라 요양급여비용을 요양기관에 지급한다.

여기서 요양급여(보험급여)란 건강보험을 통해 지급되는 가장 기본적인 급여로 건강보험 가입자 및 피부양자가 요양기관(의료기관으로 의원·병원 등)을 이용하면서 받게 되는 보험 혜택을 말한다. 건강보험의 가입자 및 피부양자는 질병·부상·출산 등에 대하여 진찰·검사, 약제·치료재료의 지급, 처치·수술·그 밖의 치료, 예방·재활 및 입원 등에 대해 요양급여를 받을 수 있다. 요양급여비용은 가입자 및 피부

1) 건강보험심사평가원(<http://www.hira.or.kr>)

양자의 질병·부상·출산 등에 대하여 요양기관으로부터 진찰·검사, 약제·치료재료의 지급, 처치·수술 기타의 치료, 예방·재활, 입원, 간호, 이송 등 의료서비스를 제공받는데 발생하는 금액을 말한다.

심사평가원에서 요양급여 등의 적정성 평가는 의약학적 측면과 비용 효과적 측면에서 요양급여를 적정하게 하였는지를 평가하고 있으며, 요양기관별·진료과목별 또는 상병별로 구분하여 평가하고 있다.

□ 수가제도

우리나라 건강보험제도는 도입 당시부터 행위별 수가제도(fee-for-service)를 실시하고 있는데, 이는 요양기관에서 의료인이 제공한 의료서비스(행위, 약제, 치료재료 등)에 대해 서비스별로 가격(수가)을 정하여 사용량과 가격에 의해 진료비를 지불하는 제도이다. 그리고 행위별 수가제의 보완 및 의료자원의 효율적 활용을 위하여 일부 질병에 대해서는 질병군별 포괄수가제(diagnosis related group, DRG)와 정액수가제(요양병원, 보건기관 등)도 병행하여 실시하고 있다.

포괄수가제는 입원환자의 진료비를 보상하는 제도로서 입원환자의 치료에 필요한 의료행위, 치료재료, 약제비용에 대해 의료서비스의 종류나 양에 관계없이 어떤 질병의 진료를 위해 입원했었는가에 따라 미리 책정된 일정액의 진료비를 보상하는 제도이다. 즉 입원비가 하나로 묶여 있다고 생각할 수 있는데, 현재는 백내장수술, 맹장수술, 항문수술, 편도수술, 탈장수술, 자궁수술, 제왕절개분만 7가지의 수술에 적용되고 있다. 7개 질병군으로 입원한 환자의 수술과 관련된 진료뿐 아니라 수술로 인한 합병증이나 환자가 입원 당시 같이 앓고 있던 질병의 치료까지 모두 보험급여 범위에 포함된다.

□ 급여와 비급여

건강보험제도는 보험자가 피보험자인 가입자로부터 보험료를 받아 보험재정을 관리하며, 가입자가 요양기관에서 진료를 받을 경우 이에 대하여 보험급여(요양기관에 대한 진료비용 지급)를 실시하는 보험자-피보험자-요양기관의 삼각관계로 이루어져 있다. 보험급여와 비급여는 진료에 대한 비용부담방식에 관한 사항으로, 관련 전문평가위원회에서 보험급여 여부와 가격을 평가하고 가입자 단체 대표, 의료공급자 대표, 여러 분야의 전문가, 정부, 보험자 등으로 구성된 위원회의 의결을 거쳐 보건복지부장관이 최종으로 고시한다.

급여는 건강보험의 적용을 받는 가입자 및 피부양자의 질병·부상에 대한 예방·진단·치료·재활과 출산·사망 및 건강증진에 대하여 국민건강보험법의 규정에 따

라 실시하는 의료서비스 또는 현금을 의미한다.

비급여는 국민건강보험법에 의하여 요양급여의 대상에서 제외되는 사항으로 국민건강보험 요양급여의 기준에 관한 규칙 제9조의 비급여 대상에 해당되는 경우를 의미하므로 비급여 환자 또는 비급여 약제에 대하여는 요양기관(또는 약국)의 관행수가(일반수가)로 본인이 전액 부담해야 한다. 일반적으로 의학적 근거, 질병 검사와 치료에 해당되지만 건강보험 재정문제로 공단에서 지불해 주지 않는 항목이라고 할 수 있다. 예를 들어 예방접종, 건강검진, 보약, 쌍꺼풀수술, 인공수정, 치아 교정치료 등이 비급여로 분류되어 있다.

비급여는 ① 제증명수수료, ② 제도적 비급여, ③ 항목별 비급여, ④ 급여기준에 의한 비급여로 크게 4종류로 분류할 수 있다. 이 중 ① 제증명수수료는 일반진단서, 장애진단서, 입원확인서 등에 해당되고, ② 제도적 비급여는 「국민건강보험 요양급여의 기준에 관한 규칙」의 (별표2) 비급여 대상에서 정한 비급여항목에 해당된다. ③ 항목별 비급여는 보건복지부 고시로 의료 행위, 치료재료 목록에서 비급여로 정한 것이고, ④ 급여기준에 의한 비급여는 급여 항목이다. 예를 들어 상급병실료 차액, 치과 임플란트, 시력교정술, 한방물리치료 등은 제도적 비급여, 다빈치로봇수술료 및 체온열검사 등은 항목별 비급여, 초음파검사, MRI 등은 급여기준에 의한 비급여에 해당된다.

따라서 비급여 환자 또는 비급여 약제에 대해 요양기관(또는 약국)의 관행수가(일반수가)로 본인이 전액 부담하여야 하므로 요양기관(또는 약국)에서 별도로 심사평가원에 비용을 청구할 수 없다.

백분의 백(100대 100) 본인부담은 「국민건강보험법」 시행규칙 제16조(요양급여비용의 본인부담) 및 동법 시행령 별표 2 제4호에 의거하여, 보험급여가 되는 진료사항을 보험수가로 계산한 전액을 본인이 부담하는 경우를 의미한다. 요양급여절차를 무시하고 요양급여기관을 이용한 비용, 해외출국 등 급여정지자의 보험진료 비용, 3개월 이상 보험료 체납시의 진료비용, 법정 요양급여일수 상한일 이후의 진료비용 등의 경우에 보험수가로 산정한 전체비용을 본인이 부담한다.

□ 의료급여 진료비심사

의료급여제도는 생활유지 능력이 없거나 생활이 어려운 저소득 국민의 의료문제(질병, 부상, 출산 등)를 국가가 보장하는 공공부조 제도로 함께 국민 의료보장의 중요한 수단인 되는 사회보장제도이다. 의료급여비용의 심사도 심사평가원이 담당하고 있는데, 의료급여 수급권자는 다음과 같이 구분할 수 있다.

<표 1-1> 의료급여 수급권자 구분

구분	대상자	
1종 수급권자	국민기초생활보장수급자	근로무능력가구, 희귀난치성질환등록자, 중증질환자, 시설수급자
	타법적용자	이재민, 의사자 및 의사자의 유족, 입양아동(18세미만), 국가유공자, 중요무형문화재 보유자, 북한이탈주민 5·18민주화운동 관련자, 노숙인
	행려환자	
2종 수급권자	국민기초생활보장 대상자 중 1종 수급대상이 아닌 가구	

□ 보훈진료비심사

심사평가원에서는 보훈국비환자(국비진료대상자)의 보훈병원 및 위탁병원의 진료비 심사·평가, 위탁병원 의사의 처방전에 의한 약국 약제비 심사·평가를 수행하고 있다. 국비진료대상자란 독립유공자예우에 관한 법률 및 국가유공자 등 예우 및 지원에 관한 법률에 따라 국가가 진료에 소요되는 비용을 부담하는 진료대상자를 말한다. 위탁병원의 경우 2005년 10월부터, 보훈병원의 경우 2008년 10월부터 심사평가원에서 국비진료대상자의 진료비에 대해 수탁심사를 담당하고 있다. 보훈진료비 심사 절차 및 방법은 원칙적으로 건강보험과 동일하다.

1.2. 진료비 청구 및 심사

심사평가원은 의료공급자가 진료비를 청구하면 국민건강보험법 등에서 정한 기준에 의해 진료비와 진료 내역이 올바르게 청구되었는지, 의·약학적으로 타당하고 비용 효과적으로 이루어졌는지 확인한다.

□ 진료비 청구

요양급여비용 청구방법, 심사청구서·명세서서식 및 작성요령(보건복지부 고시)에는 국민건강보험법 시행규칙 제19조제3항에 따라 요양급여비용의 청구방법, 요양급여비용심사청구서·요양급여비용명세서 서식과 작성요령에 관한 사항이 규정되어 있다.

요양기관이 요양급여비용을 청구하고자 하는 때(또는 처방전만을 발행하는 때)에는 요양급여비용심사청구서에 요양급여비용명세서(이하 명세서)와 기타 필요한 서류를 첨부하여 심사평가원에 제출한다. 이때 명세서를 작성할 때는 아래 제8조에 따라 입원의 경우 입원진료기간의 요양급여내역을 동일한 명세서에 통합하여 작성하며, 외래의 경우 방문일자별로 각각 작성한다.

제8조(명세서의 구분 및 작성방법)

- ① 동일 수진자에 대한 명세서는 제12조 제1항에 따라 구분하여 다음 각호와 같이 작성한다.
1. 입원의 경우 입원진료기간의 요양급여내역을 동일한 명세서에 통합하여 작성한다. 다만, 요양병원형 수가를 적용하는 수진자의 명세서는 정액수가 적용기간과 행위별수가 적용기간(특정기간)의 요양급여내역을 구분하여 각각 작성한다.
 2. 외래의 경우 요양급여내역을 방문일자별로 각각 작성한다.
 3. 약국(한국희귀의약품센타를 포함한다. 이하 같다)의 처방조제인 경우에는 처방전별로 명세서를 각각 작성하며, 직접조제인 경우에는 방문일자별로 각각 작성한다.

30일을 초과하여 입원진료를 하는 경우는 연계 심사가 가능하도록 청구구분에 분리청구코드, 기청구명세서의 접수번호, 명세서일련번호 및 최초입원개시일을 기재하도록 되어 있다.

제9조(요양급여비용 청구 및 자료제출 시기)

- ④ 30일을 초과하여 입원진료를 하는 경우에는 연계 심사가 가능하도록 다음 각호의 기재방법에 따라 월단위로 분할 청구할 수 있다. 다만, 의약품관리료의 경우에는 일괄하여 청구한다.
1. 정보통신망과 전산매체의 경우 청구구분란에 분리청구코드, 기청구명세서의 접수번호, 명세서일련번호 및 최초입원개시일을 기재한다.
 2. 서면의 경우 명세서 상단에 최초입원개시일을 기재하고, 진료내역 하단의 “특정내역”란에 기청구명세서의 접수번호, 명세서일련번호, 당월요양개시일, 수술명 및 주요진료내용, 월별 요양급여비용 총액 1 등을 기재한다.

의과입원 및 의과외래 일자별 청구를 위한 요양급여비용명세서는 다음과 같다.

(별지 제10호 서식)

의과
임원

서식번호		G I O 2				요양급여비용명세서										요양기관			
등록번호						기관번호		공상 등 구분		기호		명칭							
가 입 자 성명						분류기호		수술	진료	상해	특정	연하	환자	당월요양개시일	당월요양급여일수 (투약일수포함)	진료결과			
수진자성명														최초입원개시일		일			
상 병 명																일			
																일			
																일			
입원일수						진료행위 (I)		진료행위 (II)		처방전발달번호				처방일수					
구분																			
1. 진료료 (의과임의로 포함)	①초진	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②재진	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③의약품진료		원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	④응급 및 회송료		원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
2. 입원료	①입원	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②내과외과 등 정실환자	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③중환자실	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	④격리병실	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	⑤신생아	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	⑥기타	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	⑦기본의대	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	⑧가산의대	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
3. 투약료 및 처방전	①내복	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②외용	일	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③처방전	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
4. 주사료	①피하주사	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②정맥주사	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③정맥주사	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	④정맥주사	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	⑤정맥주사	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	⑥정맥주사	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
5. 마취료	①	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
6. 이환요령료	①	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
7. 정신요령료	①	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
8. 처치 및 수술료	①처치	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②수술	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
9. 검사료	①진단	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②진단	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진단	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
10. 영상진단 및 방사선 치료료	①진단	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②진단	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진단	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
S. 특수장비	①CT	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②MRI	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③PET	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
A. 100분의100미만 본인부담 1	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
B. 100분의100미만 본인부담 2	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
U. 건강보험100분의 100본인부담	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
V. 보훈 등 100분의100 본인부담	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
W. 의료여	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
U. 건강보험100분의 100본인부담	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
V. 보훈 등 100분의100 본인부담	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
W. 의료여	①의약품	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②치료재료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③진료행위	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
특수장비충액				원	19. 청구액				원										
보훈 등 100분의100본인부담금충액				원	20. 본인부담상한액초과금				원										
의료여충액				원	21. 요양급여비용충액 2, 진료비충액				원										
11. 소계		원		원	22. 보훈청구액				원										
12. 가산율		%		원	23. 건강보험 100분의100본인부담금충액				원										
15. 요양급여비용충액 1				원	24. 보훈본인일부부담금				원										
16. 본인일부부담금				원	25. 100분의100미만 충액				원							감 I			
17. 지원금				원	26. 100분의100미만 본인일부부담금				원							감 II			
18. 장애인의료비				원	27. 100분의100미만 청구액				원							충 I			
일련번호		※심사조정		원	28. 100분의100미만 보훈청구액				원							충 II			

[그림 1-1] 의과임원 요양급여비용명세서

(별지 제11호 서식)

의과
외래

서식번호		G	I	O	3	요양급여비용명세서										요양기관			
등록번호						환원호								기호					
가입자성명						공상등구분								명칭					
수진자성명																			
상		별		명		분류기호	수술	치료	진료	특정	진료	진료	진료	진료	내원일자	요양급여일수 (원내부담일수포함)	진료결과		
구		분		기본진료 약제,특정 재료(I)	진료행위 (II)	처방전 발달번호	진료일수	처방일수	본인부담금 발생횟수	진료 횟수	진료 횟수	진료 횟수	진료 횟수	진료 횟수	진료 횟수	진료 횟수	진료 횟수		
1.진료 (외과관리료 포함)	①초진	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	②재진	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	③의약품 ④진료 ⑤진료 ⑥진료 ⑦진료 ⑧진료 ⑨진료 ⑩진료 ⑪진료 ⑫진료 ⑬진료 ⑭진료 ⑮진료 ⑯진료 ⑰진료 ⑱진료 ⑲진료 ⑳진료 ㉑진료 ㉒진료 ㉓진료 ㉔진료 ㉕진료 ㉖진료 ㉗진료 ㉘진료 ㉙진료 ㉚진료 ㉛진료 ㉜진료 ㉝진료 ㉞진료 ㉟진료 ㊱진료 ㊲진료 ㊳진료 ㊴진료 ㊵진료 ㊶진료 ㊷진료 ㊸진료 ㊹진료 ㊺진료 ㊻진료 ㊼진료 ㊽진료 ㊾진료 ㊿진료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
	①초진	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
3.투약료 및 처방전	①내부 ②외부 ③처방전	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
4.주사료	①파라 ②정맥 ③정맥 ④정맥 ⑤정맥 ⑥정맥 ⑦정맥 ⑧정맥 ⑨정맥 ⑩정맥 ⑪정맥 ⑫정맥 ⑬정맥 ⑭정맥 ⑮정맥 ⑯정맥 ⑰정맥 ⑱정맥 ㉑정맥 ㉒정맥 ㉓정맥 ㉔정맥 ㉕정맥 ㉖정맥 ㉗정맥 ㉘정맥 ㉙정맥 ㉚정맥 ㉛정맥 ㉜정맥 ㉝정맥 ㉞정맥 ㉟정맥 ㊱정맥 ㊲정맥 ㊳정맥 ㊴정맥 ㊵정맥 ㊶정맥 ㊷정맥 ㊸정맥 ㊹정맥 ㊺정맥 ㊻정맥 ㊼정맥 ㊽정맥 ㊾정맥 ㊿정맥	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
5.마취료	①	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
6.이환요법료	①	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
7.정신요법료	①	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
8.치치 및 수술료	①치치 ②치치 ③치치 ④치치 ⑤치치 ⑥치치 ⑦치치 ⑧치치 ⑨치치 ⑩치치 ⑪치치 ⑫치치 ⑬치치 ⑭치치 ⑮치치 ⑯치치 ⑰치치 ⑱치치 ㉑치치 ㉒치치 ㉓치치 ㉔치치 ㉕치치 ㉖치치 ㉗치치 ㉘치치 ㉙치치 ㉚치치 ㉛치치 ㉜치치 ㉝치치 ㉞치치 ㉟치치 ㊱치치 ㊲치치 ㊳치치 ㊴치치 ㊵치치 ㊶치치 ㊷치치 ㊸치치 ㊹치치 ㊺치치 ㊻치치 ㊼치치 ㊽치치 ㊾치치 ㊿치치	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
9.검사료	①자 ②자 ③자 ④자 ⑤자 ⑥자 ⑦자 ⑧자 ⑨자 ⑩자 ⑪자 ⑫자 ⑬자 ⑭자 ⑮자 ⑯자 ⑰자 ⑱자 ㉑자 ㉒자 ㉓자 ㉔자 ㉕자 ㉖자 ㉗자 ㉘자 ㉙자 ㉚자 ㉛자 ㉜자 ㉝자 ㉞자 ㉟자 ㊱자 ㊲자 ㊳자 ㊴자 ㊵자 ㊶자 ㊷자 ㊸자 ㊹자 ㊺자 ㊻자 ㊼자 ㊽자 ㊾자 ㊿자	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
10.영상진단 및 영상진 치료료	①진 ②진 ③진 ④진 ⑤진 ⑥진 ⑦진 ⑧진 ⑨진 ⑩진 ⑪진 ⑫진 ⑬진 ⑭진 ⑮진 ⑯진 ⑰진 ⑱진 ㉑진 ㉒진 ㉓진 ㉔진 ㉕진 ㉖진 ㉗진 ㉘진 ㉙진 ㉚진 ㉛진 ㉜진 ㉝진 ㉞진 ㉟진 ㊱진 ㊲진 ㊳진 ㊴진 ㊵진 ㊶진 ㊷진 ㊸진 ㊹진 ㊺진 ㊻진 ㊼진 ㊽진 ㊾진 ㊿진	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
5.특수장비	①CT ②MRI ③PET	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
T.특수재료 및 관련 행위료	①치료 ②치료 ③치료 ④치료 ⑤치료 ⑥치료 ⑦치료 ⑧치료 ⑨치료 ⑩치료 ⑪치료 ⑫치료 ⑬치료 ⑭치료 ⑮치료 ⑯치료 ⑰치료 ⑱치료 ㉑치료 ㉒치료 ㉓치료 ㉔치료 ㉕치료 ㉖치료 ㉗치료 ㉘치료 ㉙치료 ㉚치료 ㉛치료 ㉜치료 ㉝치료 ㉞치료 ㉟치료 ㊱치료 ㊲치료 ㊳치료 ㊴치료 ㊵치료 ㊶치료 ㊷치료 ㊸치료 ㊹치료 ㊺치료 ㊻치료 ㊼치료 ㊽치료 ㊾치료 ㊿치료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
A.100분의100미만 본인부담1	①의약품 ②치료 ③진료 ④진료 ⑤진료 ⑥진료 ⑦진료 ⑧진료 ⑨진료 ⑩진료 ⑪진료 ⑫진료 ⑬진료 ⑭진료 ⑮진료 ⑯진료 ⑰진료 ⑱진료 ㉑진료 ㉒진료 ㉓진료 ㉔진료 ㉕진료 ㉖진료 ㉗진료 ㉘진료 ㉙진료 ㉚진료 ㉛진료 ㉜진료 ㉝진료 ㉞진료 ㉟진료 ㊱진료 ㊲진료 ㊳진료 ㊴진료 ㊵진료 ㊶진료 ㊷진료 ㊸진료 ㊹진료 ㊺진료 ㊻진료 ㊼진료 ㊽진료 ㊾진료 ㊿진료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
B.100분의100미만 본인부담2	①의약품 ②치료 ③진료 ④진료 ⑤진료 ⑥진료 ⑦진료 ⑧진료 ⑨진료 ⑩진료 ⑪진료 ⑫진료 ⑬진료 ⑭진료 ⑮진료 ⑯진료 ⑰진료 ⑱진료 ㉑진료 ㉒진료 ㉓진료 ㉔진료 ㉕진료 ㉖진료 ㉗진료 ㉘진료 ㉙진료 ㉚진료 ㉛진료 ㉜진료 ㉝진료 ㉞진료 ㉟진료 ㊱진료 ㊲진료 ㊳진료 ㊴진료 ㊵진료 ㊶진료 ㊷진료 ㊸진료 ㊹진료 ㊺진료 ㊻진료 ㊼진료 ㊽진료 ㊾진료 ㊿진료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
U.건강보험100분의 100본인부담	①의약품 ②치료 ③진료 ④진료 ⑤진료 ⑥진료 ⑦진료 ⑧진료 ⑨진료 ⑩진료 ⑪진료 ⑫진료 ⑬진료 ⑭진료 ⑮진료 ⑯진료 ⑰진료 ⑱진료 ㉑진료 ㉒진료 ㉓진료 ㉔진료 ㉕진료 ㉖진료 ㉗진료 ㉘진료 ㉙진료 ㉚진료 ㉛진료 ㉜진료 ㉝진료 ㉞진료 ㉟진료 ㊱진료 ㊲진료 ㊳진료 ㊴진료 ㊵진료 ㊶진료 ㊷진료 ㊸진료 ㊹진료 ㊺진료 ㊻진료 ㊼진료 ㊽진료 ㊾진료 ㊿진료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
V.보훈등 100분의100 본인부담	①의약품 ②치료 ③진료 ④진료 ⑤진료 ⑥진료 ⑦진료 ⑧진료 ⑨진료 ⑩진료 ⑪진료 ⑫진료 ⑬진료 ⑭진료 ⑮진료 ⑯진료 ⑰진료 ⑱진료 ㉑진료 ㉒진료 ㉓진료 ㉔진료 ㉕진료 ㉖진료 ㉗진료 ㉘진료 ㉙진료 ㉚진료 ㉛진료 ㉜진료 ㉝진료 ㉞진료 ㉟진료 ㊱진료 ㊲진료 ㊳진료 ㊴진료 ㊵진료 ㊶진료 ㊷진료 ㊸진료 ㊹진료 ㊺진료 ㊻진료 ㊼진료 ㊽진료 ㊾진료 ㊿진료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
W.의급여	①의약품 ②치료 ③진료 ④진료 ⑤진료 ⑥진료 ⑦진료 ⑧진료 ⑨진료 ⑩진료 ⑪진료 ⑫진료 ⑬진료 ⑭진료 ⑮진료 ⑯진료 ⑰진료 ⑱진료 ㉑진료 ㉒진료 ㉓진료 ㉔진료 ㉕진료 ㉖진료 ㉗진료 ㉘진료 ㉙진료 ㉚진료 ㉛진료 ㉜진료 ㉝진료 ㉞진료 ㉟진료 ㊱진료 ㊲진료 ㊳진료 ㊴진료 ㊵진료 ㊶진료 ㊷진료 ㊸진료 ㊹진료 ㊺진료 ㊻진료 ㊼진료 ㊽진료 ㊾진료 ㊿진료	회	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	
특약료총액			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
특수장비총액			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
특수재료 및 관련 행위료 총액			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
보훈 등 100분의100 본인부담금 총액			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
의급여총액			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
11. 소개		원	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
12. 가산율		%	원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
15. 요양급여비용총액 1			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
16. 본인일부부담금			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
일련번호			원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		
수진자 일련번호	-		원	원	원	원	원	원	원	원	원	원	원	원	원	원	원		

190mm×320mm(일반용지 60g/㎡(재활용품))

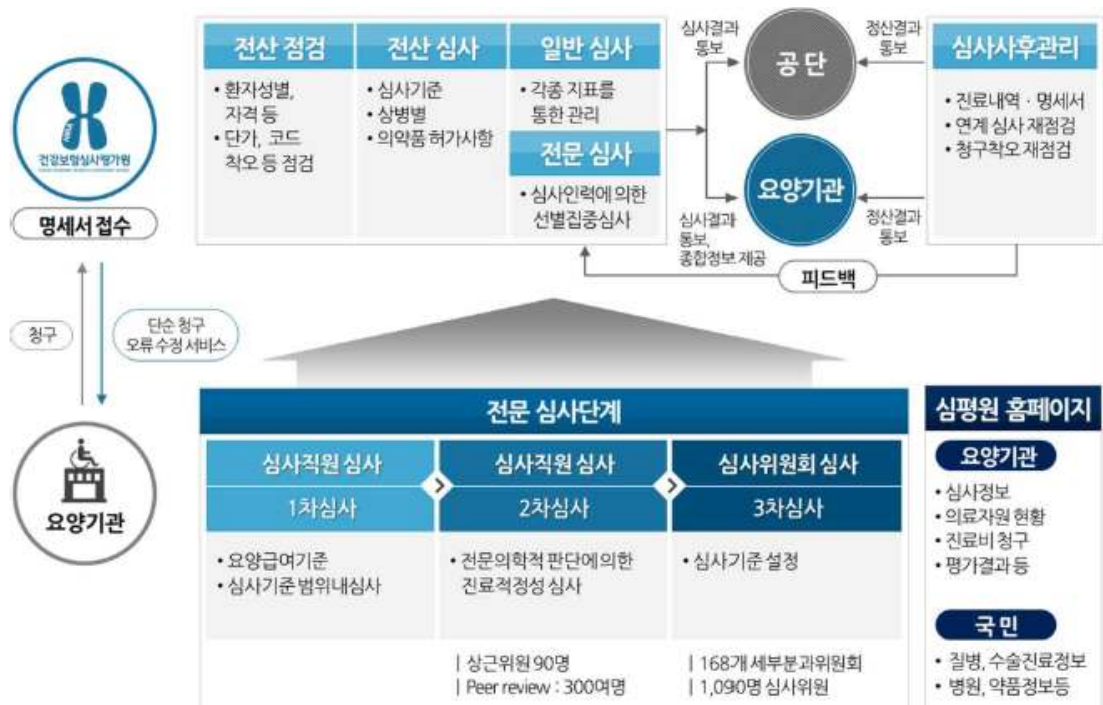
[일 자 별]

[그림 1-2] 의과외래 요양급여비용명세서

□ 진료비 심사

먼저 진료비심사처리 절차를 살펴보면, 진료비 심사는 의료공급자가 비용청구 양식인 ‘진료비 청구명세서’에 환자별로 진료한 내역과 비용을 청구하면서부터 시작된다.

- ① [청구명세서접수] 의료공급자가 심사평가원의 진료비청구프로그램을 이용하여 청구 전에 청구파일을 점검하고 인터넷 망을 통해 심사평가원에 직접 청구한다.
- ② [전산점검] 모든 청구명세서의 환자 성별 및 자격, 환자 상병코드, 청구코드 및 가격의 오류 등에 대한 점검이 전산프로그램에 의해 이루어진다.
- ③ [인공지능 전산심사] IT 기술과 심사직원의 심사 노하우를 접목하여 사람이 심사하는 것과 같이 로직화된 전산프로그램으로, 인공지능 자동화심사를 말하며, 모든 진료내역은 7단계 전산심사가 이루어진다. 이때 급여기준, 의약품 허가사항 등에 대한 심사가 이루어진다.
- ④ [전문심사] 착오청구 개연성이 높거나 전문의학적 판단이 필요한 건을 심사자가 직접 심사하는 것으로, 일차적으로 심사직원이 의료공급자의 청구경향을 파악한 후 청구방법, 산정지침 및 심사기준의 적합성 여부 등을 심사한다. 그리고 전문의학적 판단에 의한 진료적정성 심사로 해당분야 전문의사가 진료 기록의 확인 등 보완자료를 요청하거나 진료의사와의 면담 및 방문을 통해 심사하는 심사위원 심사를 실시한다. 마지막으로 심사위원회를 개최하여 전문 과목별로 전문 의학적 판단에 의한 새로운 기준이 필요한 사항, 심사기준 설정을 요하는 사항, 심사기준 적용에 이견이 있는 사항, 기타 합의에 의한 결정을 필요로 하는 사항 등을 여러 전문가가 모여서 심사하는 심사위원회 심사가 있다.
- ⑤ [심사결과 통보] 전문 심사가 완료되면, 진료비 지급 업무를 담당하고 있는 국민건강보험공단에 심사결과를 통보하고, 동시에 진료비를 청구한 요양기관에도 심사결과를 통보한다.
- ⑥ [심사 사후관리] 심사가 완료된 건 중 수진자별, 진료기간별 또는 의료공급자 간 연계가 되지 않아 미처 급여기준 적용하지 못한 것에 대하여 추가적인 심사를 하여 지급된 비용을 환수한다. 의료공급자와 국민건강보험공단은 심사평가원의 심사결정에 대하여 수용할 수 없다고 판단되는 경우에 관련 자료를 첨부하여 이의 신청을 할 수 있다.



[그림 1-3] 진료비심사처리 절차

(출처 : 빅데이터 분석을 통한 질병 감시에측 서비스)

진료비심사 시스템에서 진료비 심사가 완료된 데이터는 매일 DW시스템으로 자동 전송되며 전송된 데이터는 변환, 정제과정을 거쳐 EDW(Enterprise Data Warehouse), Data Mart, Summary Tables로 관리된다. DW시스템에 적재된 데이터는 코드 값 등에 대한 오류 점검을 통해 데이터의 정합성과 신뢰성을 확보하게 된다.

1.3. 보건의료빅데이터개방시스템²⁾

□ 자료제공 체계

심사평가원은 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 근거하여 보건의료빅데이터개방시스템을 통하여 보건의료빅데이터인 의료정보를 편리하고 손쉽게 활용할 수 있도록 국민에게 개방하고 있다. 보건의료빅데이터개방시스템 홈페이지에서 제공하는 공공데이터는 누구나 이용 가능하고, 영리 목적의 이용을 포함한 자유로운 활용이 보장된다. 여기서 의료정보란 의료기관 청구자료 기반의 자료로 전 국민의 진료정보, 의약품 정보, 치료재료정보, 의료자원정보 등을 분석·정제한 데이터를 말한다.

2) 보건의료빅데이터개방시스템(<http://opendata.hira.or.kr>)

○ 자료제공대상

- 연구자료
 - 국가, 지방자치단체 및 정부 산하기관
 - 연구 중심 병원 및 학술 연구 수행기관 등 연구목적으로 신청하는 자
- 산업체자료
 - 의약품 품목 허가를 받은 자
 - 제약, 의료기기 산업
 - 컨설팅 회사 및 예비 창업자 등

○ 자료제공범위

- 보건의료 자료 제공 시 건강보험자료 범위 내에서 제공
- 맞춤형 자료 제공 시 건강보험심사평가원에서 제공 가능한 형태의 자료 제공

○ 자료제공기준

- 공공데이터에 개인정보 등 비공개대상정보가 포함된 경우 (공공기관의 정보공개에 관한 법률) 제공 제한
- 주민등록번호, 성명, 주소, 전화번호, 요양기관기호(명칭)등 개인정보와 개별 법인·단체 등의 정보가 식별 불가능한 형태로 제공
- 개인정보 등 비공개대상정보를 기술적으로 분리할 수 있는 경우 개인정보 등 비공개대상정보를 제외한 공공데이터는 부분제공
- 공공데이터의 정보가 다른 정보와 결합하여 개인 식별이 가능한 경우 제공 제한

□ 제공자료 내역

심사평가원에서 제공하는 자료는 진료정보, 의약품정보, 치료재료정보, 의료자원정보로 크게 구분할 수 있으며, 진료정보에는 전 국민의 의료이용이 모두 포함되어 있는 건강보험 청구자료(이하 ‘청구자료’)가 포함되어 있다.

<표 1-2> 제공자료 내역

구분	주요 정보	DB명칭
진료정보	청구명세서 정보	청구명세서 DB
	의료행위 정의 및 환자 분류 정보	정보분석 DB
	수가마스터 정보	통계자료 DB
	질병군(DRG)·요양병원 수가 마스터정보	질병통계 DB
	의료행위별 심사기준 정보(보장 범위) 의료행위 18개 분류별 진료규모 정보(진료량, 금액) 질병군(DRG) 및 요양병원 진료규모 정보 질병정보(주상병) 및 질병단위 진료규모 정보	행위통계 DB
의약품 정보	급여의약품 마스터 정보 급여의약품 사용 정보	의약품DB
치료재료 정보	치료재료 마스터 정보 치료재료별 사용정보 특수 재료관련 정보(복강경시술, 조영제 등)	치료재료 DB
의료자원 정보	요양기관 개폐업 정보	요양기관현황 DB
	의료기관 시설 정보(병상, 집중치료실, 수술실 등) 인력(의사, 간호사, 의료기사 등) 현황 장비 보유현황(장비별 이력관리)	요양기관 종합정보 DB

□ 자료제공 방법

심사평가원에서는 환자데이터셋 또는 보건의료빅데이터개방시스템을 통해 자료를 제공하고 있다.

○ 환자데이터셋

환자데이터셋은 청구자료를 기초로 하여 진료개시일 기준 1년간 진료 받은 환자를 대상으로 표본을 추출한 자료이다. 환자데이터셋은 원자료에서 개인 및 법인에 대한 정보를 제거한 후, 1년간 의료서비스를 이용한 모든 환자를 대상으로 진료내역과 처방내역을 포함하여 성별, 연령구간(5세 단위)에 따른 환자단위 층화계통 추출한 자료로 연도별, 표본자료 종류별, 환자, 요양기관 매칭이 불가능하다. 제공되고 있는 환자데이터셋의 종류는 다음 표와 같으며, 개방시스템 홈페이지에서 신청가능하다.

<표 1-3> 환자데이터셋의 종류

구분	설명
입원환자데이터셋(HIRA-NIS)	- 2009년, 2010년, 2011년, 2012년, 2013년, 2014년, 2015년 데이터 제공 - 입원환자 추출비율 13% (약 70만 명) - 외래환자 추출비율 1% (약 40만 명)
전체환자데이터셋(HIRA-NPS)	- 2010년, 2011년, 2012년, 2013년, 2014년, 2015년 데이터 제공 - 전체환자 추출비율 3% (약 140만 명)
고령환자데이터셋(HIRA-APS)	- 2010년, 2011년, 2012년, 2013년, 2014년, 2015년 데이터 제공 - 고령환자(65세 이상) 추출비율 20% (약 100만 명)
소아청소년환자데이터셋(HIRA-PPS)	- 2010년, 2011년, 2012년, 2013년, 2014년, 2015년 데이터 제공 - 소아청소년환자(20세 미만) 추출비율 10% (약 110만 명)

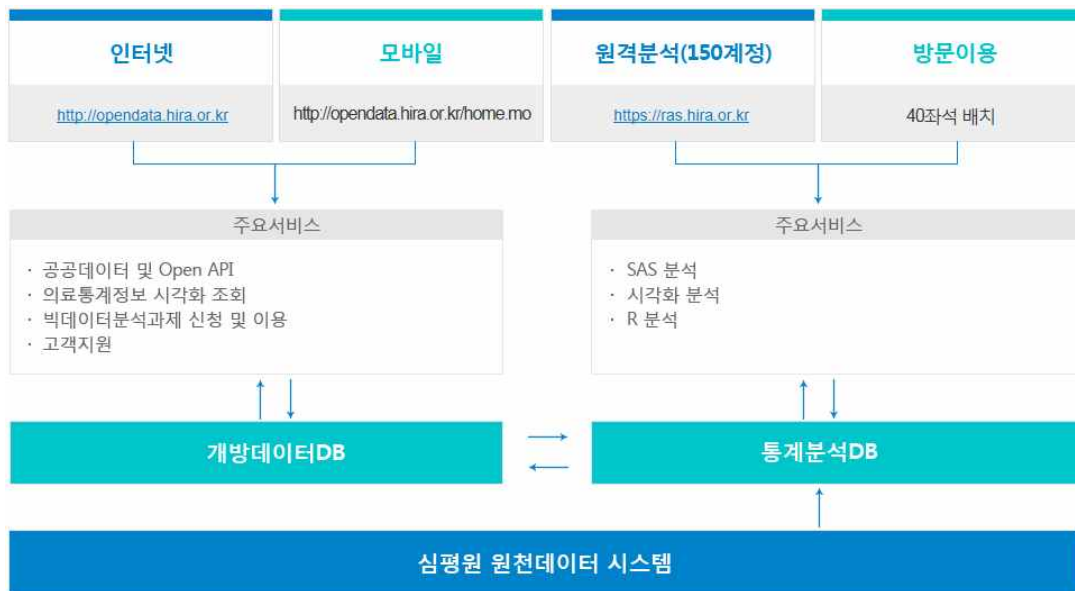
○ 보건의료빅데이터개방시스템

심사평가원의 보건의료빅데이터개방시스템은 1) 개방데이터 DB를 이용한 인터넷 및 모바일을 통한 데이터 및 통계자료 조회 서비스와 2) 통계분석 DB를 이용한 원격 접속시스템 및 보건의료빅데이터센터(이하 ‘빅데이터센터’) 운영으로 구성되어 있다.

원격접속시스템은 이용자가 원하는 장소에서 원격접속(가상화PC)을 통한 분석지원 서비스로 주로 유관기관 및 학계의 연구목적의 자료분석 등을 위해 이용되고 있으며, 총 150계정이 운용되고 있다(2016년12월 기준). 빅데이터센터는 유관기관 및 학계의 연구목적의 자료분석 및 의료계 및 산업계의 R&D 개발을 위한 공간으로, 본원 및 전국에 총 40석을 운영하고 있다.

산업계자료는 심평원에서 운영하고 있는 빅데이터센터에 방문하여 분석을 수행하고 있으며, 연구과제는 IRB 심의 통보서, 기관 간 자료요청 공문, 보안각서, 보안준수 협약서 등의 정보보안을 준수하는 조건으로 원격분석 지원을 이용하고 있다. 원격통계분석시스템 접속 시에는 이용자 ID, 공인IP, 사설IP, MAC 주소, 공인인증서 등을 통해 승인된 이용자만 접속 가능하다.

제공데이터는 온라인, 전자매체 등의 방법으로 제공되며, 개인정보 보호를 위한 대체키 변환 등을 통해 제공된다.



[그림 1-4] 보건의료빅데이터개방시스템 구성

□ 자료신청단계

심사평가원에 자료신청을 하는 단계는 다음 그림과 같다. 먼저 이메일을 통해 상담신청을 하면, 심사평가원에서 상담신청내용을 검토하며, 검토가 끝나면 신청자가 홈페이지에서 이용신청을 한다. 상담신청부터 홈페이지에 이용신청을 하기까지 대략 10일 소요되며, 홈페이지에 신청을 완료하면 심사평가원에서는 자료요청을 접수하고 신청한 자료를 추출하여 원격시스템을 통해 제공한다. 자료 신청부터 데이터분석 자료를 원격시스템을 통해 제공하기까지 보통 2~4주 정도 소요된다. 자세한 자료신청 절차는 개방시스템 홈페이지에 별도로 안내되어 있는 가이드를 참고한다.



[그림 1-5] 자료신청 단계

공공데이터 제공 수수료 산정기준은 공공데이터의 제공에 소요되는 인건비 등 실비와 공공데이터의 제공을 위하여 필요한 정보시스템 및 네트워크 증설·유지보수 비용, 데이터의 양, 제공기간 및 횟수를 고려하여 정해진다.

□ 개인정보/민감정보 관련 정책 및 법적 근거

「국민건강보험법」 시행령 제81조에는 심사평가원이 민감정보 및 고유식별정보를 다룰 수 있는 경우에 대해 다음과 같이 명시하고 있다. 여기서 제63조제1항에 따른 업무는 요양급여비용의 심사, 요양급여의 적정성 평가, 심사기준 및 평가기준의 개발, 이에 따른 업무와 관련된 조사연구 및 국제협력, 다른 법률에 따라 지급되는 급여비용의 심사 또는 의료의 적정성 평가에 관하여 위탁받은 업무, 건강보험과 관련하여 보건복지부장관이 필요하다고 인정한 업무, 그 밖에 보험급여 비용의 심사와 보험급여의 적정성 평가와 관련하여 대통령령으로 정하는 업무를 말한다. 또한 심사평가원은 「국민건강보험법」 제96조에 따라 요양급여비용을 심사하고 요양급여의 적정성을 평가하기 위하여 주민등록 등의 자료를 제공하도록 요청할 수 있다.

국민건강보험법 시행령 제81조(민감정보 및 고유식별정보의 처리)

② 심사평가원은 다음 각 호의 사무를 수행하기 위하여 불가피한 경우 「개인정보 보호법」 제23조에 따른 건강에 관한 정보, 같은 법 시행령 제19조 각 호에 따른 주민등록번호, 여권번호, 운전면허의 면허번호 또는 외국인등록번호가 포함된 자료를 처리할 수 있다.

1. 법 제48조에 따른 요양급여 대상 여부의 확인 등에 관한 사무
2. 법 제63조제1항에 따른 업무에 관한 사무
3. 법 제87조 및 제90조에 따른 이의신청 및 행정소송에 관한 사무
4. 법 제96조에 따른 자료의 제공 요청에 관한 사무

국민건강보험법 제96조(자료의 제공)

② 심사평가원은 국가, 지방자치단체, 요양기관, 「보험업법」에 따른 보험회사 및 보험료를 산출 기관, 「공공기관의 운영에 관한 법률」에 따른 공공기관, 그 밖의 공공단체 등에 대하여 요양급여비용을 심사하고 요양급여의 적정성을 평가하기 위하여 주민등록·출입국관리·진료기록·의약품공급 등의 자료로서 대통령령으로 정하는 자료를 제공하도록 요청할 수 있다.

따라서 심사평가원은 국민건강보험법에 명시된 고유 업무 수행을 위해 개인정보 및 민감정보를 처리할 수 있는 법적 근거가 있다고 할 수 있다.

□ 자료제공 관련 정책 및 법적 근거

심사평가원은 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따라 보유·관리하고 있는 공공데이터를 국민에게 제공하고 있다. 심사평가원에서는 기본적으로 비식별화된 자료를 제공하고 있으나, 외부자료원을 청구자료와 연계하여 제공할 필요가 있는 경우도 있다. 따라서 비식별화된 자료를 제공하는 경우와 자료연계가 필요한 경우 자료제공과 관련된 법적 근거는 다음과 같다.

○ 비식별화된 자료제공

심사평가원은 개인을 알아볼 수 없는 형태로 비식별화된 자료를 별도의 DB로 구축하고 빅데이터센터를 통해 접속하여 통계작성 및 학술연구 등에 활용하도록 하고 있다. 따라서 「개인정보 보호법」 제18조제2항제4호의 ‘통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우’에 해당된다고 할 수 있다. 이용자는 비식별화된 데이터의 열람 및 분석만 가능하고 자료를 내려 받는 것은 원천 차단되며, 분석 결과인 통계값에 한하여 사전 승인 후 반출 가능하다.

환자의 의료정보의 권리는 환자와 의료인에게 있는데, 제3자의 경제적 이익을 도모하기 위해 제공하는 것이 권리 침해 행위가 아닌지에 대한 우려도 있는데, 심사평가원의 데이터 개방은 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따른 것으로 동법 제3조제4항에는 ‘공공데이터의 영리적 이용인 경우에도 이를 금지하거나 제한하여서는 아니 된다’고 규정하고 있다.

○ 자료연계

외부기관의 자료원을 개인의 고유식별정보를 이용하여 심사평가원의 청구자료와 연계하여 분석을 수행하고자 할 경우, 외부기관이 민감정보, 고유식별정보, 주민등록번호를 심사평가원에 제공하기 위해서는 제3자에 제공에 대한 법적 근거가 있어야 한다. 「개인정보보호법」과 「생명윤리 및 안전에 관한 법률」을 바탕으로 살펴보면, 서면동의를 받고, 기관위원회의 심의를 받은 경우 민감정보, 고유식별정보, 주민등록번호 처리가 가능하고, 제3자에게 제공이 가능하다. 따라서 환자를 대상으로 수집한 자료를 청구자료와 연계하여 분석을 하고자 할 경우, 외부기관에서는 ① 환자들을 대상으로 민감정보, 고유식별정보, 주민등록번호 활용, 제3자에 제공에 대한 별도 서면 동의를 취득해야 한다. 그리고 ② 기관위원회의 심의를 거쳐 승인을 받아야 한다. 이 경우 외부기관은 환자 자료와 주민등록번호를 심사평가원에 반입하여 주민등록번호를 기준으로 자료연계를 시도할 수 있다.

<표 1-4> 자료연계 관련 법적 근거

법률	비고
개인정보보호법	<ul style="list-style-type: none"> - 개인정보의 제3자 제공 : 동의를 받은 경우 - 민감정보, 고유식별정보, 주민등록번호 처리가 가능한 경우 ·별도로 동의를 받은 경우 ·법령에서 구체적으로 고유식별 정보의 처리를 요구하거나 허용하는 경우
생명윤리 및 안전에 관한 법률	<ul style="list-style-type: none"> - 제3자에게 제공 ·서면동의를 받고 기관위원회의 심의를 거친 경우 ·익명화가 원칙이나 개인식별 정보 포함에 대한 동의를 한 경우 익명화 하지 않아도 됨 - 기관위원회의 승인을 받아 서면동의를 면제할 수 있는 요건(둘다 만족) ·연구대상자의 동의를 받는 것이 연구 진행과정에서 현실적으로 불가능하거나 연구의 타당성에 심각한 영향을 미친다고 판단되는 경우 ·연구대상자의 동의 거부를 추정할 만한 사유가 없고, 동의를 면제하여도 연구대상자에게 미치는 위험이 극히 낮은 경우

만일 민감정보, 고유식별정보, 주민등록번호 활용, 제3자에 제공에 대한 별도 서면 동의서를 받지 않은 경우에는 「생명윤리 및 안전에 관한 법률」 제16조에 따라 다음 두 가지 요소를 모두 갖춘 경우 기관위원회의 승인을 받아 연구대상자의 서면동의를 면제할 수 있다.

1. 연구대상자의 동의를 받는 것이 연구 진행과정에서 현실적으로 불가능하거나 연구의 타당성에 심각한 영향을 미친다고 판단되는 경우
2. 연구대상자의 동의 거부를 추정할 만한 사유가 없고, 동의를 면제하여도 연구대상자에게 미치는 위험이 극히 낮은 경우

○ 환자표본자료

심사평가원의 환자데이터셋은 1년간 의료이용을 한 모든 환자를 대상으로 통계학적으로 120만명을 추출하여 표본추출 환자의 1년간 진료·처방 내역에 대한 자료로서, 「통계법」 제3조제4호에 따른 통계자료에 해당된다. 여기서 ‘통계자료’란 통계작성기관이 통계의 작성을 위하여 수집·취득 또는 사용한 자료(데이터베이스 등 전산자료 포함)를 의미한다.

환자표본자료 제공에 대한 법적 근거는 「통계법」 제31조 및 같은 법 시행령 제47조부터 제49조까지의 규정으로 통계자료에 관한 「공공기관의 정보공개에 관한 법률」의 특례 규정에 해당한다. 따라서 개인정보가 삭제되었고 표본집단 이므로 어떠한 조합으로 생성된 개인이라 할지라도 실제 모집단의 동일조합의 환자 수는 알 수 없으므로, 표본자료는 특정 개인이나 법인 또는 단체 등을 식별할 수 없는 형태라고 할 수 있으므로, 「통계법」 제31조에 따라 제공가능하다고 할 수 있다.

통계법 제31조(통계자료의 이용)

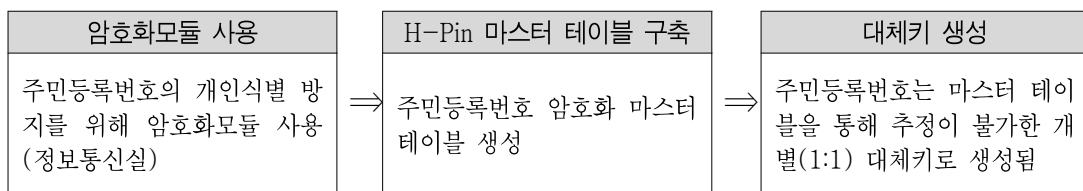
①통계자료를 이용하고자 하는 자는 대통령령으로 정하는 바에 따라 통계작성기관의 장에게 통계자료의 제공을 신청할 수 있다.

②통계작성기관의 장은 제1항에 따른 신청을 받은 때에는 통계자료의 사용목적·내용 및 범위의 타당성을 심사하여 타당하다고 판단되고, 영업상 비밀을 침해할 가능성이 없는 경우에는 이를 제공하여야 한다. 이 경우 통계작성기관의 장은 다음 각 호의 경우를 제외하고는 특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다.

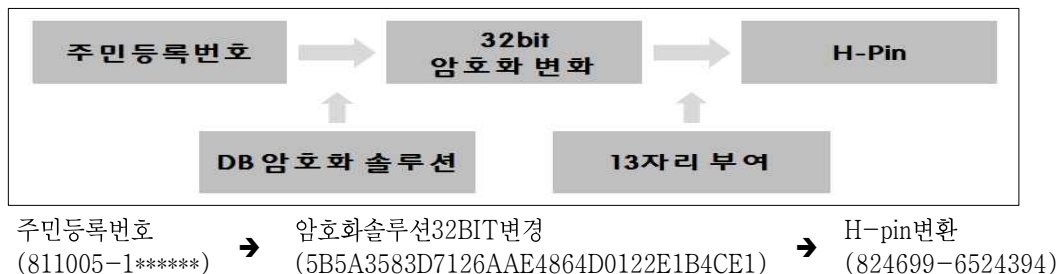
1. 통계응답자가 자신이 응답한 자료를 요구하는 경우
2. 총조사 및 제18조에 따라 통계청장의 승인을 받아 작성하는 통계 중 사업체를 대상으로 하는 전수조사를 통하여 취득한 정보로서 사업체 명, 업종, 주소 등 대통령령으로 정하는 정보를 제공하는 경우

□ 빅데이터 개방 DB 정보보호 처리

심사평가원에서는 빅데이터 개방 DB의 정보보호 처리를 위하여 두 단계에 걸쳐 정보보호를 실시하고 있다. 먼저 1차 정보보호 단계에서 원천 데이터를 개인식별 정보를 제외한 후 개방 DB에 구축하여 활용하며, 임의의 대체키(H-Pin) 변환을 통해 개인정보의 추정 자체가 원천적으로 불가능하도록 한다. 이때 국가정보원 CC(Common Criteria 국제공통평가기준)인증을 획득한 암호화모듈을 사용하고 있으며, 여기서 개인식별 정보는 주민자의 성명, 주민등록번호 등 개인식별 가능 정보를 말한다.



H-PIN 변환과정을 구체적으로 살펴보면, 먼저 주민등록번호를 DB암호화솔루션을 이용하여 32bit 암호화변화를 수행하고, 13자리의 H-Pin으로 다음과 같이 변환한다.



2차 정보보호 단계에서는 임의의 일련번호(H-PIN), 요양기관기호, 명세서번호 등

고유식별키를 이중 비식별처리한다. 즉 자료 활용 및 자료제공 시 변환된 H-PIN을 임의의 난수로 다시 한 번 대체 처리를 한다.



추가로 요양기관기호와 명세서번호는 제공·활용 과정에서 이중 비식별화 처리하는데, 자료 제공 시점마다 일련번호, 요양기관기호, 명세서번호는 임의의 난수로 변환 처리한다. 그리고 이용자는 구축한 DB에 접속하여 열람 및 분석만 가능하며, 자료를 내려 받는 것은 원천 차단되고 분석 결과 값에 대해서는 승인 후 반출 가능하도록 하고 있다.

○ 영업정보 마스킹

제약업체의 영업정보 보호를 위하여 의약품코드는 마스킹 처리하고 일반명코드를 제공하고 있으며, 치료재료업체의 영업정보 보호를 위하여 업체코드인 세분류 부분을 마스킹하여 제공하고 있다.

1.4. 건강보험 청구자료의 구성³⁾

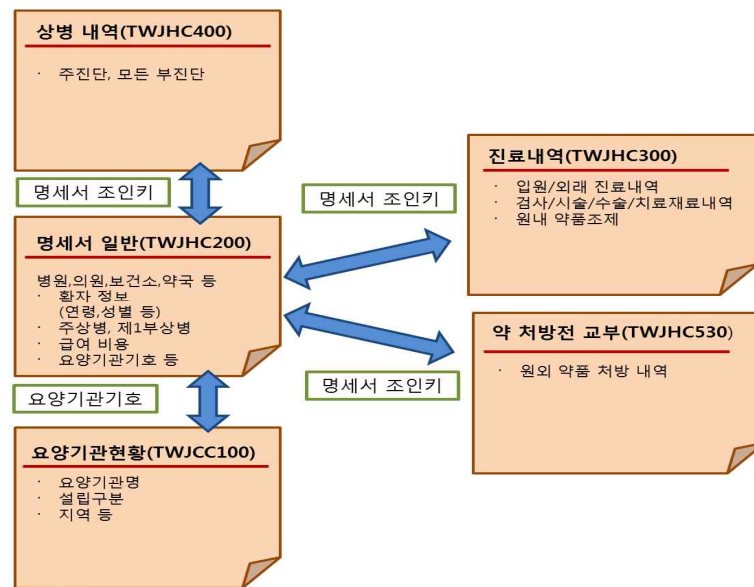
‘건강보험 청구자료(이하 청구자료)’란 요양기관에서 환자의 진료비용 중 ‘국민건강보험’이 부담하는 부분에 대해 지급의뢰를 하기 위해 심사평가원에 보험급여청구를 하면서 발생하는 정보를 말한다.

청구자료의 장점으로 1) 요양기관을 방문한 환자의 비급여를 제외한 모든 의료서비스 이용 정보가 포함되어 분석결과의 일반화가 가능하고, 2) 의료서비스 내역(시술/수술, 검사, 약제처방, 치료재료 등), 서비스별 급여비용 등이 포함되어, 환자가 받은 개별적인 의료서비스를 확인 가능하며, 3) 환자 고유 식별자가 자료에 포함되어 환자의 추적 자료 구축이 가능하다는 점이다.

청구자료는 여러 테이블로 구성되어 있으며, 이 중 연구에 많이 활용되는 테이블은 명세서일반내역 테이블(이하 T200), 진료내역 테이블(이하 T300), 수진자상병내역 테이블(이하 T400), 원외처방전상세내역(이하 T530), 요양기관테이블 등이다. 요양기관으로부터 명세서가 청구되면, 명세서에 포함된 내용들을 분류하여 각 테이블에 저장한다. 청구된 한 명세서에 대해 T200 테이블에는 한 줄로 저장되며, T200 테이블에는 각 명세서를 구분하는 명세서 조인키가 부여되어 있다. T200 테이블의 명세서 조인키는 중복키가 없으며, 한건의 청구 건을 나타낸다. T200 테이블의 한 건의 명세

3) 건강보험 청구자료 분석 매뉴얼. 건강보험심사평가원. 2015.

서 조인키에 대해 검사, 시술, 수술, 치료재료, 원내 약품처방에 대한 세부진료내역이 T300 테이블에 저장되어 있으며, 주상병 및 모든 부상병 정보는 T400 테이블에 저장되어 있다. 의료기관에서 원외처방한 약품 정보는 T530 테이블에 저장되어 있으며, T300, T400 및 T530 테이블의 경우 T200 테이블의 한 건의 명세서에 대해 여러 줄로 정보가 저장되어 있다. 추가적으로 요양기관 현황에 대한 정보를 이용하기 위해 서는 요양기관기호를 기준으로 T200 테이블과 연결하여 사용할 수 있다.



[그림 1-6] 건강보험 청구자료 테이블 구조

각 테이블의 세부 조사내용은 다음 표와 같다.

<표 1-5> 건강보험 청구자료 테이블 구성

한글테이블명	영문테이블명	주요 조사내용
명세서일반내역 테이블	T200(TWJHC200)	환자 및 기관의 기본정보, 기본적인 상병(주상병명과 첫 번째 부상병명), 내원일수, 요양개시일자, 특정기호구분 등의 일반적 현황
진료내역 테이블	T300(TWJHC300)	요양기관 내에서의 처치, 수술, 원내 약처방, 청구금액 등 상세 의료서비스 현황
수진자상병내역 테이블	T400(TWJHC400)	T200의 주상병과 부상병을 포함한 추가 부상병명, 일련번호 '1'은 T200의 주상병과 동일하며, 일련번호 '2'는 T200의 부상병과 동일함
원외처방전상세내역 테이블	T530(TWJHC530)	원외로 처방되는 약에 대한 정보가 있는 테이블(원내처방내역은 T300에 포함되어 있음). 처방된 약마다 한 줄로 구성 되어 KEY값 변수를 기준으로 동일명세서는 여러 줄로 되어있고 줄번호 변수가 있음. 이때 줄번호는 T300에 있는 줄번호와 다름

요양기관 테이블	YKHO(TWJHC100)	요양기관의 지역, 종별, 설립구분 등 요양기관에 대한 일반적 정보
-------------	----------------	---

□ 원격시스템 제공 자료

유관기관 및 학계의 연구목적의 자료분석 등을 위해 이용되고 있는 원격시스템에서 제공하고 있는 테이블과 변수는 다음과 같다.

<표 1-6> 명세서일반내역(T200)

한글변수명	영문변수명	변수유형	변수 크기	변수설명
명세서조인키	SPEC_ID	BIGINT	8	명세서 레벨의 팩트간 조인키(각 자료 간 연결 키):접수번호+접수년도+지원코드+청일년+명일련(고유키로 중복 없음)
보험자코드	INSUP_TP_CD	CHAR	1	1:직장, 3:지역, 4:건강보험, 5:의료급여, 6:지역3차, 7:국비보훈(상이처, 무자격자), 9:무료진료
수진자개인 식별대체키	PAT_HPIN	BIGINT	8	수진자당 부여된 일련번호 or 주민등록번호 대체번호(임의의 연번 부여) 가상의 개인 식별자 (외국인, 무료진료, 정 보기관 소속된 사람은 주민번호가 없음, 신 생아의 경우 출생 신고 전에는 주민번호가 없으므로 개인식별자 별도요청이 필요함)
성별구분	SEX_TP_CD	CHAR	1	수진자 수급자의 성별구분 1:남, 2:여, 9:기타
수진자 연령	PAT_AGE	NUM	4	청구시점에서의 연령 0.2:28일미만, 0.9:28일이상 1세미만, 999:연령 산출 불가, 5세단위로 범주화하여 제공
수진자통계 연령	PAT_STC_AGE	NUM	4	매년 말일 기준의 수진자 연령 0.2:28일미만, 0.9:28일이상 1세미만, 999:연령 산출 불가('진료연도-태어난 연도'이며, 동일연도는 모두 같은 '만 나이'를 가짐)
요양기관식별 대체키	YID	BIGINT	8	요양기관기호 대체번호, 가상의 요양기관 식 별자(월별 별도의 임의의 연번 부여)
요양종별코드	CL_CD	CHAR	2	01:상급종합병원, 11:종합병원, 21:병원, 28:요 양병원, 29:정신요양병원, 31:의원, 41:치과병 원, 51:치과의원, 61:조산원, 71:보건소, 72:보 건지소, 73:보건진료소, 74:모자보건센터, 75: 보건의료원, 81:약국, 92:한방병원, 93:한의원

한글변수명	영문변수명	변수유형	변수 크기	변수설명
지역코드	RVD_PLC_CD	CHAR	6	요양기관 소재한 지역(시, 군, 구)의 코드, 앞 2자리가 시도코드임 [17개시도] 11:서울 21:부산 22:인천 23:대구 24:광주 25: 대전 26:울산 31:경기 32:강원 33:충북 34:충 남 35:전북 36:전남 37:경북 38:경남 39:제주, 41:세종 (요양기관 소재지, 더 세부적인 지역구분이 필요한 경우별도 요청)
서식구분코드	FOM_TP_CD	CHAR	3	[서식코드를 기준으로 보건기관의 의과, 치 과, 한방으로 세분화] 021:의과입원, 031:의과외래, 041:치과입원, 051:치과외래, 061:조산원입원, 071:보건기관 입원의과, 072:보건기관입원치과, 073:보건기 관입원한방, 081:보건기관외래의과, 082:보건 기관외래치과, 083:보건기관외래한방, 091:정 신과낮병동, 101:정신과입원, 111:정신과외래, 121:한방입원, 131: 한방외래, 151:의료급여혈 액투석, 201:직접조제, 211:처방조제, 991: 조 산원외래
주상병코드	MAIN_SICK	CHAR	7	주된 상병분류기호: 약국 증상분류기호, 한 방상병코드, 의과:표준질병사인분류코드(KCD 코드, 상병명을 3자리만 고려하면 ‘상세불 명’이 포함되므로 자료요청 전 상세 불명 포함할지 여부 검토할 것),(배제진단 제외)
부상병코드	SUB_SICK	CHAR	7	주된 상병분류기호 외의 추가 상병분류기호
진료과목코드	DGSBJT_CD	CHAR	2	요양기관에서 청구시 기재한 진료과목코드, 의원급 표시과목코드 00:일반의, 01:내과, 02:신경과, 03:정신과, 04:외과, 05:정형외과, 06:신경외과, 07:흉부외 과, 08:성형외과, 09:마취통증의학과, 10:산부 인과, 11:소아청소년과, 12:안과, 13:이비인후 과, 14:피부과, 15:비뇨기과, 16:영상의학과, 17:방사선종양학과, 18:병리과, 19:진단검사의 학과, 20:결핵과, 21:재활의학과, 22:핵의학과, 23:가정의학과, 24:응급의학과, 25:산업의학 과, 26:예방의학과, 49:치과, 50:구강악안면외 과, 51:치과보철과, 52:치과교정과, 53:소아치 과, 54:치주과, 55:치과보존과, 56:구강내과, 57:구강악안면방사선과, 58:구강병리과, 59:예 방치과, 80:한방내과, 81:한방부인과, 82:한방 소아과, 83:한방안·이비인후·피부과, 84:한 방신경정신과, 85:침구과, 86:한방재활의학과, 87:사상체질과, 88:한방응급, 89:한방응급, 90: 한방소계, 99:기타, (\$:진료과목오류, AA:개설 과목전체, ZZ:ZZZZZZ)

한글변수명	영문변수명	변수유형	변수 크기	변수설명
요양개시일자	RECU_FR_DD	CHAR	8	수진자가 진료를 받기 시작한 일자, 조제투여일자, 최초내방일자(청구명세서 단위)입원인 경우 해당 명세서의 입원진료년월외래인 경우 내원한 진료년월
요양종료일자	RECU_TO_DD	CHAR	8	수진자가 진료 받기를 종료한 일자
최초입원일자	FST_DD	CHAR	8	30일을 초과하여 계속 입원한 환자의 요양급여비용을 분리청구하는 경우, 해당 입원의 최초 입원일자(년/월/일)를 기재함(추가청구분의 경우는 안함),(2004. 7 심사분부터, NULL이면 요양개시일자와 동일)
입내원일수	VST_DDCNT	NUM	3	수진자가 진료를 받기 위해 요양기관에 내원한 일수(초진+재진). 입원일수(외래명세서 일자별 청구 : 2010년부터 시작, 2012년 전국 확대)
요양일수	RECU_DDCNT	NUM	3	수진자를 진료한 총일수(투약 일수 포함)
원외처방일수	OPRSC_DDCNT	NUM	9	처방전 발행 내역 중 처방일수의 합계
원외처방약제비	OPRSC_DGMAMT_AMT	NUM	18	처방전 발행 내역 중 약제비 금액의 합계
원외처방전건수	OPRSCD_CNT	NUM	9	처방전을 발행한 건수
청구요양급여비용총액	DMD_RPE_TAMT_AMT	NUM	8	요양기관에서 수진자 진료에 사용한 건강보험 총 요양 급여비용(100/100 본인 부담금 및 비급여 비용 제외) (심사이전 청구된 금액)
청구본인부담금	DMD_SLF_BRDN_AMT	NUM	8	건강보험 총요양급여비용 중 본인이 부담해야할 금액 (심사이전 청구된 금액)
청구보험자부담금	DMD_INSUP_BRDN_AMT	NUM	8	수진자가 요양기관에서 진료 받은 총요양급여비용 중 보험자가 부담해야할 금액 (심사이전 청구된 금액)
심사결정요양급여비용총액	RVD_RPE_TAMT_AMT	NUM	18	심사결과 수진자 본인이 부담해야 될 금액과 보험자가 부담해야 도리 금액을 합한 요양급여비용총액 요양기관에서 수진자 진료에 사용한 건강보험 총 요양 급여비용(100/100 본인 부담금 및 비급여 비용 제외) (심사결정 후 금액)
심사결정본인부담금액	RVD_SLF_BRDN_AMT	NUM	18	건강보험 총 요양급여비용 중 본인이 부담해야할 금액 (심사결정 후 금액)
심사결정보험자부담금액	RVD_INSUP_BRDN_AMT	NUM	18	수진자가 요양기관에서 진료 받은 총 요양급여비용 중 보험자가 부담해야할 금액 (심사결정 후 금액)
수술여부	SOPR_YN	CHAR	1	주된 상병명과 관련된 수술을 시행한 경우 0:미수술, 9:수술

한글변수명	영문변수명	변수유형	변수크기	변수설명
공상구분코드	OINJ_TP_CD	CHAR	1	공무상 상해 구분 0:정상건, 1:공상건(공고공단), 3:보훈감면 30%(광주민주유공자), 4:보훈국비(건강보험 또는 의료급여 수급권자), 5:보훈감면50%, 6:보훈감면60%, 7:보훈국비환자(상이처, 무자격자), 8:군인가족, 예비역장군 및 대령등 대상(공고공단, 직장, 지역분), 9:군인,군무원 대상(공고공단분) B:보훈병원(상이처, 무자격자/보험급여1차), C:차상위 본인부담 경감대상자, D:보훈병원(보험급여2차), E:차상위 만성질환 18세미만 본인부담 경감대상자, F:차상위 장애인 만성질환 18세미만 본인부담 경감대상자 G:긴급복지 의료지원대상자 H:희귀난치성 질환 지원대상자
특정기호구분	PRCL_SYM_CD	CHAR	4	상병정보의 특정기호. 상병정보에 특정기호가 복수 개인 경우 일련번호가 가장 작은 건의 특정기호로 설정
상해외인구분 코드	INJ_EXA_TP_CD	CHAR	1	한국표준질병사인분류의 제20장에 의거 상병의 원인에 해당하는 분류기호중 영문 첫자리(V, W, X, Y)만 기재. 다만 의료급여수진자가 정신과입원진료 중 다른 진료과목 전문의의 치료를 받아 별도의 해당 진료비 명세서를 작성하는 경우 분류기호 'E'기재 [한국표준질병사인분류의 제20장에 상병의 원인에 해당되는 분류기호] (V: V상해외인, W: W상해외인, X: X상해외인, Y: Y상해외인)
진료결과구분	DGRSLT_TP_CD	CHAR	1	요양급여비용명세서상 최종 진료일의 환자 상태를 구분하여 기재 1: 계속 2: 이송 3: 회송 4: 사망 5: 기타 9: 퇴원 또는 외래 치료종결 (9번 2010년 추가. 의원급의 진료결과변수는 부정확함)
입원도착경로 구분	IPAT_ARIV_PTH_TP	CHAR	2	첫째자리:도착경로 (1:타기관경유, 2:응급구조대 후송, 3:기타) 둘째자리:입원경로 (1:응급실, 2:외래)
의료급여종별	MAIDCL_CD	CHAR	1	[의료급여 대상자의 의료급여종별 구분코드] 1:1종(생활보호대상자중 거택보호자, 사회복지시설수용), 2:2종(생활보호대상자중자활보호자), 4:4종(행여자), 5:5종(성병), 6:6종(2종장애인 2차진료-입원서식), 8:8종(2종장애인1차진료-외래서식), \$:의료급여종별코드요류
청구형태코드	DMD_FRM_CD	CHAR	1	1:서면, 2:디스켓, 3:EDI, 4:DRG (DRG, 보훈, 정신질환, 의료급여 등은 진료 세부내역이 없어 주로 EDI 만을 요청함. DRG 병의원급으로 확대예정(2012 하반기))

한글변수명	영문변수명	변수유형	변수 크기	변수설명
청구구분코드	DMD_TP_CD	CHAR	1	0:일반청구, 1:보완청구, 2:추가청구, 3:분리 청구, 4:본인부담상한액초과원청구, 5:본인부 담상한액초과보완청구, 6:본인부담상한초과 추가청구, 7:본인부담상한액초과분리청구 (4-7:2004.7.1.부터사용) 8:약제상한차액추가 청구(2010.8추가)
심사년월	RV_YM	CHAR	6	건강보험심사평가원에서 심사가 이루어진 년월

<표 1-7> 진료내역(T300)

한글변수명	영문변수명	변수유형	변수 크기	변수설명
명세서조인키	SPEC_ID	BIGINT	8	명세서 레벨의 팩트간 조인키(각 자료 간 연결 키):접수번호+접수년도+지원코드+청일년+명일련(고유키로 중복 없음)
항목코드	CZ_ITEM_CD	CHAR	5	<p>명세서 항별 목코드(항코드+목코드) 각 항코드마다 목코드 분류가 있어 항코드와 목코드를 조합하여 진료내역의 일반적 특성을 파악할 수 있음 [항코드·목코드]</p> <p>01:진찰료 (01:초진,02:재진,03:응급 및 회송료등)</p> <p>02:입원료 (01:일반,02:내과,정신과,만8세미만의소아,03:중환자실,04:격리병실,10:기본식대,11:가산식대,99:기타입원료)</p> <p>03:투약료 (01:내복약,02:외용약,03:처방전)</p> <p>04:주사료(01:주사,99:기타)</p> <p>05:마취료(01:마취)</p> <p>06:이학요법료(01:이학요법료)</p> <p>07:정신요법료(01:정신요법료)</p> <p>08:처치 및 수술료 (01:처치 및 수술,(치과)보통처치외 처치항목,02:(치과)절개외 수술항목03:캐스트,99:치과기타)</p> <p>09:검사료(01:자체검사,02:위탁검사)</p> <p>10:영상진단 및 방사선치료료 (01:진단, 02:치료)</p> <p>L:요양병원정액(01:요양병원 정액수가)</p> <p>S: 특수장비 (01:CT진단,02:MRI진단,03:PET진단)</p> <p>V:보훈 등100/100본인부담 (01:의약품,02:치료재료,03:진료행위)</p> <p>W:비급여(*한방 경우는 다름) (01:의약품,02:치료재료,03:진료행위)</p> <p>X:정신과 정액 (01:의료급여 정신질환 정액수가(외래, 낮병동, 입원, 외박수가), 02:투약 1일당 정액수가)</p>

한글변수명	영문변수명	변수유형	변수 크기	변수설명
분류코드구분	DIV_TY_CD	CHAR	9	[명세서의 분류 코드 유형] 0:각목소계, 1:수가, 2:준용, 3:국산보험등재약, 4:수입약/원료약/조제(제제)약, 5: 약가의 일반명(성분처방), 7:협약재료, 8:치료재료(일반재료) A:한방수가, B:한방준용수가, C:한방약가, G:한방구협약재료, H:한방치료재료, X:약국수가, Y:요양기관조제제약(5.성분처방은 제품에 따라 약가가 모두 다르므로 평균 금액을 적용함)
분류코드	DIV_CD	CHAR	9	수가(행위)코드, 약품코드, 재료대코드 등 진료내역코드 기재 (2010년 1월 이전:EDI코드/ 2010년1월 이후:KD코드이므로, 2010년 전후 자료 이용 시 두 가지 코드 모두 확인필요)
1회투약량	FQ1_MDCT_QTY	NUM	9	1회 환자에게 투약한 량
1일투약량	DY1_MDCT_QTY	NUM	9	1일 환자에게 투약한 량
1일투여량 또는 실시횟수	DY1_INJC_QTY_EX EC_FQ	NUM	18	1일 실시횟수(소수 셋째자리에서 4사5입하여 소수 둘째자리까지)을 기재, 의약품 및 처방 내역 의약품의 경우 1일 투여횟수 기재
총투여일수 또는 실시횟수	TOT_INJC_DDCNT_ EXEC_FQ	NUM	9	총 투여일수 또는 실시횟수
총사용량 또는 실시횟수	TOT_USE_QTY_OR_ EXEC_FQ	NUM	18	1회투약량 × 1일 투여(실시)횟수 × 총투여일수 또는 실시횟수
단가	UNPRC	NUM	18	「약제 및 치료재료의 비용에 대한 결정기준」에 따른 단가 분류코드별 단가(상한가)
금액	AMT	NUM	18	단가 × (일일투여량 또는 실시횟수) × (총투여일수 또는 실시횟수)=금액
가산적용금액	ADDC_ADT_AMT	NUM	18	2란 행위 가산이 있는 경우 요양기관 종별 가산율을 적용한 금액 가산율을 적용한 금액
일반명코드(주 성분코드)	GNL_CD	CHAR	9	약가 주성분 등재약, 수입약의 일반명 코드
진료예외구분 코드	EXP_TP_CD	CHAR	2	의약분업 예외 사항에 해당되어 원내 조제 투약이 이루어진 경우 예외구분 코드를 기재 00:해당사항없음, 01:약국이 없는 지역, 재해 발생지역, 11:응급환자, 13:정신보건법에 의한 정신질환자, 15: 전염병예방법에 의한 제 1종 전염병환자, 29:협진(한양방, 양한방, 양양방)환자, 52:주사제를 원내 투약한 경우 (2001.11.15. 신설), 53:항암제중 주사제, 55:검사, 수술 및 처치에 사용하는 의약품, 61: 군인, 경찰, 소방공무원에 대하여 조제하는 경우, 99: 퇴장방지의약품의 사용장려 비용을 청구하는 경우

한글변수명	영문변수명	변수유형	변수 크기	변수설명
1_2구분	SP1_SP2_TP_CD	CHAR	1	진료코드의 영양기관 종별 가산율 적용여부를 구분하기 위한 코드 I : 기본진료료, 약제, 치료재료, 위탁검사 및 혈액료 등 상대가치점수표상의 영양기관 종별가산율이 적용되지 아니한 항목 II : 영양기관 종별가산율이 적용되는 진료행위 항목 1:가산없음, 2:가산있음

<표 1-8> 수진자상병내역(T400)

한글변수명	영문변수명	변수유형	변수크기	변수설명
명세서조인키	SPEC_ID	BIGINT	8	명세서 레벨의 팩트간 조인키(각 자료 간 연결 키):접수번호+접수년도+지원코드+청일년+명일련(고유키로 중복 없음)
일련번호	SICK_SNO	NUM	2	한국표준질병사인분류의 상병분류기호
상병코드	SICK_CD	CHAR	7	통계청고시 제1993-3호(1993.11.20.)에 의거 한국 표준 질병사인분류의 상병분류기호 (약국 증상분류기호, 한방상병코드, 의과:표준질병사인 분류코드) 자릿수1 - A:서식이 보건기관이고 진료과가 한방이 아닌 경우이거나 B, C가 아닌경우 B:서식이 보건기관이고 진료과가 한방인 경우이거나 서식이 한방인 경우 C:서식이 직접조제인 경우 \$:상병기호가 "이거나 마스터에 코드가 없는 경우 발생 자릿수2-6 - 상병분류기호(약국 증상분류기호, 한방상병코드, 의과:표준질병사인분류코드)
상병분류구분	SICK_TY_CD	CHAR	1	1:주상병 2:부상병 3:배제진단 (2007.10.1.추가) (2007.10.1.종합병원급 대상 신설, 2011.1.1. KCD-6차 적용)
청구진료과목 코드	DMD_DGSBJT_CD	CHAR	2	진료를 받은 진료과목(병원급이상) 또는 상병명에 해당하는 진료과목(의원급)을 기재하되, 진료과목이 2개 이상에 해당되는 경우 상병별로 모두 기재 01:내과, 02:신경과, 03:정신과, 04:외과, 05:정형외과, 06:신경외과, 07:흉부외과, 08:성형외과, 09:마취통증의학과, 10:산부인과
청구요양개시 일자	RECU_FR_DD	CHAR	8	요양기관에 그 상병 진료를 위하여 그달에 최초 내원한 연월일을 기재 단, 수진자의 요양급여비용명세서가 동일 청구서에서 정액, 정률 등으로 명세서가 분리되는 경우는 해당 요양급여비용명세서의 최초 진료일자를 기재

한글변수명	영문변수명	변수유형	변수크기	변수설명
내과세부전문 과목	IFLD_DTL_SPC_SBJ T_CD	CHAR	2	내과 진료과목 중 '세부전문의 제도인증 규명(대한의학회)'에 따라 인증 받은 세부 전문과목을 운영하고 있는 종합병원, 상급종합 병원의 경우 진료를 받은 세부전문과목 00:내과통합(요양기관에서 내과 세부전문과 목을 운영하지 않는 경우)(2007.10.1. 종합병원 원금 대상 신설) 01:소화기내과, 02:순환기내과, 03:호흡기내 과, 04:내분비.대사내과, 05:신장내과, 06:혈 액종양내과, 07:감염내과, 08:알레르기내과, 09:류마티스내과, 11:소아감염, 12:소아내분비

<표 1-9> 원외처방전내역(T530)

한글변수명	영문변수명	변수유형	변수크기	변수설명
명세서조인키	SPEC_ID_SNO	BIGINT	8	명세서 레벨의 팩트간 조인키(각 자료 간 연결 키):접수번호+접수년도+지원코드+청일년+명일련(고유키로 중복 없음)
처방전 교부번호	PRSCP_GRANT_NO	CHAR	13	진료기관에서 처방전 발행시 부여한 교부 처방전교부일자+일련번호
처방전 일련번호	PRSCP_SNO			진료기관에서 처방전 발행시 부여한 교부번호에 대한 일련번호
줄번호 (일련번호)	LN_NO	NUM	8	진료내역 줄번호로 항, 목 순으로 연이어 부여
분류코드구분	DIV_TY_CD	CHAR	1	명세서의 분류 코드 구분 (1:수가 3:약가 4:원료/조제/제제 7:협약재 8: 치료재료)
분류코드	DIV_CD	CHAR	9	약품코드 기재 (2010년 1월 이전:EDI코드/ 2010년1월 이 후:KD코드이므로, 2010년 전후 자료 이용 시 두 가지 코드 모두 확인필요)
1회 투약량	FQ1_MDCT_QTY	NUM	9	1회 환자에게 투약한 량
1일 투여횟수	DY1_MDCT_QTY	NUM	9	1일 환자에게 투약횟수
총 투여일수	TOT_INJC_DDCNT_ EXEC_FQ	NUM	9	총투여일수 또는 실시횟수
총사용량	TOT_USE_QTY_OR_ EXEC_FQ	NUM	18	1회투약량 × 1일투여횟수 × 총투약일수 = 총사용량
단가	UNPRC	NUM	18	「약제 및 치료재료의 비용에 대한 결정기준」에 따른 단가 분류코드별 단가(상한가)
금액	AMT	NUM	18	단가 × 1회 투약량 × (1일투여(실시)횟수) × (총투여일수 또는 실시횟수) = 금액

<표 1-10> 요양기관(YKHO)

한글변수명	영문변수명	변수유형	변수 크기	변수설명
요양기관대체 키	YID	CHAR	20	요양기관기호 대체번호, 가상의 요양기관 식 별자(월별 별도의 임의의 연번 부여)
최종청구월	DATE	CHAR	6	최종청구월
시도코드	SIDO_CD	CHAR	2	요양기관 소재한 지역(시, 군, 구)의 코드, 앞 2자리가 시도코드임 [17개시도] 11:서울 21:부산 22:인천 23:대구 24:광주 25: 대전 26:울산 31:경기 32:강원 33:충북 34:충 남 35:전북 36:전남 37:경북 38:경남 39:제주, 41:세종 (요양기관 소재지, 더 세부적인 지역구분이 필요한 경우별도 요청)
종별코드	CL_CD	CHAR	2	01:상급종합병원, 11:종합병원, 21:병원, 28:요 양병원, 29:정신요양병원, 31:의원, 41:치과병 원, 51:치과의원, 61:조산원, 71:보건소, 72:보 건지소, 73:보건진료소, 74:모자보건센터, 75: 보건의료원, 81:약국, 92:한방병원, 93:한의원
설립구분	ORG_DF	CHAR	2	설립구분
병상규모	BED_GRADE	NUM	2	\$:해당사항 없음, 01:0~29병상, 02:30~99병상, 03:100~299병상, 04:300~399병상, 05:400~499 병상, 06:500~999병상, 07:1000병상 이상
50병상당 의사수	PER_50BED_DOCTO R	NUM	8	
50병상당 치과의사수	PER_50BED_DENTA L	NUM	8	
50병상당 한의사수	PER_50BED_ORIEN TAL	NUM	8	
50병상당 간호사수	PER_50BED_NURSE	NUM	8	
CT보유여부	CT	NUM	8	0=무, 1=유
MRI보유여부	MRI	NUM	8	0=무, 1=유
PET보유여부	PET	NUM	8	0=무, 1=유

□ 산업계 제공 자료

의료계 및 산업계의 R&D 개발을 위한 공간으로 이용되고 있는 빅데이터센터를 이용하는 경우 제공하고 있는 변수는 다음과 같다.

<표 1-11> 산업계 변수 테이블

한글변수명	영문변수명	변수유형	변수 크기	변수설명
명세서일련번호	MID	CHAR	20	명세서당 부여된 일련번호(임의의 연번 부여)
요양개시년월	RECU_FR_YM	CHAR	6	외래인 경우 내원한 진료년월 입원인 경우 해당 명세서의 입원진료년월
심사년월	RV_YM	CHAR	6	건강보험심사평가원에서 심사가 발생한 년월
처방진료구분코드	PRSC_DIAG_TP_CD	CHAR	1	처방(C), 진료(J) 구분코드
DW주상병코드	DW_MSICK_CD	CHAR	7	양방(A) + 상병코드 or 한방(B) + 상병코드를 나타내는 주상병코드
DW부상병코드	DW_SSICK_CD	CHAR	7	양방(A) + 상병코드 or 한방(B) + 상병코드를 나타내는 부상병코드
수진자일련번호	JID	CHAR	20	수진자당 부여된 일련번호(임의의 연번 부여)
수진자연령	PAT_AGE	NUMERIC	4.1	5세 단위로 범주화하여 제공
성별	SEX_TP_CD	CHAR	1	수진자 수급자의 성별구분 1: 남, 2: 여, 9: 기타
보험자종별	INSUP_TP	CHAR	1	4: 건강보험, 5: 의료급여, 7: 보훈, 9: 무료 진료
요양기관일련번호	YID	CHAR	20	요양기관당 부여된 일련번호 (월별 별도의 임의의 연번 부여)
종별코드	L_CD	CHAR	2	01: 상급종합병원, 11: 종합병원, 21: 병원, 28: 요양병원, 29: 정신요양병원, 31: 의원, 41: 치과병원, 51: 치과의원, 61: 조산원, 71: 보건소, 72: 보건지소, 73: 보건진료소, 74: 모자보건센터, 75: 보건의료원, 81: 약국, 92: 한방병원, 93: 한의원
지역코드	PLC_CD	CHAR	6	요양기관 소재한 지역(시, 군, 구)의 코드
총병상수	TOT_SBD_CNT	INTEGER	4	요양기관의 총 병상수로 범주화하여 제공

한글변수명	영문변수명	변수유형	변수크기	변수설명
서식코드	FOM_CD	CHAR	2	021: 의과입원, 031: 의과외래, 041: 치과입원, 051: 치과외래, 061: 조산원입원, 071: 보건기관입원의과, 072: 보건기관입원치과, 073: 보건기관입원한방, 081: 보건기관외래의과, 082: 보건기관외래치과, 083: 보건기관외래한방, 091: 정신과낮병동, 101: 정신과입원, 111: 정신과외래, 121: 한방입원, 131: 한방외래, 151: 의료급여혈액투석정액, 201: 약국직접조제, 211: 약국처방조제, 991: 조산원외래
통합분류코드 (제품명)	UNL_DIV_CD	CHAR	17	주성분코드로 요청 시 일부분만 제공
약효분류코드	MEFT_DIV_NO	CHAR	3	약효가 유사한 동일 효능(약효)군
ATC코드	WHO_ATC_CD	CHAR	7	WHO에서 개발한 국제의약품 분류 코드
주성분코드	GNL_NM_CD	CHAR	9	주성분코드(일반명코드)
진료과목코드	DGSBJT_CD	CHAR	2	진료를 받은 진료과목(병원급 이상) 또는 상병명에 해당하는 진료과목(의원급)
표시과목코드	SHW_SBJT_CD	CHAR	2	요양기관이 외부에 표시하기 위해 정의한 표시과목
내과세부과목 코드	IFLD_DTL_SPC_SBJT_CD	CHAR	2	내과 진료과목 중 '세부전문의 제도인증 규정(대한의학회)'에 따라 인증받은 세부전문의과목을 운영하고 있는 종합병원, 상급종합병원의 경우 진료를 받은 세부전문의과목
내원일수	VST_DDCNT	INTEGER	4	입원 또는 내원하여 진료를 받은 실 일수
1회투약량	FQ1_MDCT_QTY	NUMERIC	18	1회투약량
1일투여횟수	DY1_INJC_QTY_EXEC_FQ	NUMERIC	18	1일투여횟수
총투여일수	TOT_INJC_DDCNT_EXEC_FQ	INTEGER	4	총 투여일수 또는 실시횟수
총사용량	TOT_USE_QTY_OR_EXEC_FQ	NUMERIC	18	1회투약량 * 1일투여횟수 * 총투여일수
금액	AMT	BIGINT	8	단가 * 1회 투약량 * 1일 투여량 (투여(실시)횟수) * 총투여일수(실시횟수)
단가	UNPRC	BIGINT	8	「약제 및 치료재료의 비용에 대한 결정기준」에 따른 단가

1.5. 자료이용 시 주의사항

청구자료를 이용하여 분석할 경우 다음과 같은 사항을 고려하여 자료의 제한점, 연구의 제한점 등을 검토하고 결과 해석시 참고해야 한다.

첫째, 요양기관에서 명세서를 청구할 때 해당 명세서의 주진단명 및 부진단명을 상병코드로 기입한다. 원칙적으로 해당 방문과 관련있는 진단명을 기입해야 하지만

기준에 입력된 상병코드가 그대로 저장되어 있거나, 급여 청구를 위해 필요한 상병코드를 기입하는 경우 등이 발생하여 상병만으로 특정 질환을 정의할 경우 질환 정의에 대한 정확도가 떨어질 수 있다. 따라서 상병명 뿐만 아니라 처치코드, 약물코드 등을 활용하여 질환 정의의 민감도를 높일 필요가 있다.

둘째, 동일 명세서 내에서의 검사, 처치, 약물처방 등의 시간적 순서를 파악할 수 없다. 예를 들어 특정한 결과 발생을 예방하기 위한 시술과 특정한 결과 발생 후 치료를 목적으로 수행되는 시술이 동일한 경우, 동일한 명세서 내에서는 해당 시술이 예방적 치료에 해당하는지, 발생 후 치료에 해당하는지를 구분할 수 없다. 따라서 이런 명세서 자료의 특성을 고려하여 연구를 설계해야한다.

셋째, 입원 명세서의 경우 동일 입원인 경우에도 분리하여 별도의 명세서로 청구되는 경우가 많은데, 동일 입원에 대한 분석이 필요한 경우 명세서의 입원 에피소드 등으로 조작적으로 정의할 필요가 있다.

넷째, 청구자료 수집 기준에 따라 포괄수가제 또는 요양기관 정액제 등과 같이 세부진료내역이 없는 명세서가 존재한다. 따라서 의료기술 및 결과변수, 교란요인 정의 시 세부진료내역을 이용하여 관련 변수 정의를 수행했을 경우 해당 변수가 과소평가될 가능성이 있다.

다섯째, 청구자료를 이용하여 변수를 정의할 경우 분류 비뚤림(classification bias)이 발생할 수 있는데, 포괄수가제 또는 급여제한 등으로 실제로 해당 변수에 대한 경험 이 있는데도 경험이 없는 것으로 오분류할 수 있으며, 약물의 경우 처방받았지만 복용하지 않았을 경우 해당 변수에 대한 경험이 없는데도 경험이 있는 것으로 오분류할 가능성이 있다.

예를 들면 스텐트 치료재료는 환자당 평생 3개만 급여가 되는 치료재료이다. 스텐트 시술을 받는 경우 스텐트 시술코드와 스텐트 치료재료 코드가 모두 청구되는 것이 일반적인데, 스텐트 시술코드와 치료재료 수가코드는 다음 표와 같다.

<표 1-12> 스텐트 관련 수가코드

수가코드		
스텐트 시술 코드	경피적 관상동맥 스텐트 삽입술	M6561,M6562,M6563,M6564
치료재료코드	금속스텐트	J5083001,J5083011,J5083016,J5083029,J8083033,J5083040,J5083046,J5083070,J5083094,J5083101,J5083106,J5083107,J5083137,J5083140,J5083173,J5083201,J5083206,J5083213,J5083240,J5083273,J5083301,J5083401,J5083501,J5084040,J5084140,J8083006,J8083073,J8083089,J8083129,J8083340,J8083373,J8083473
	약물방출스텐트	J5231024,J5231061,J5231067,J5231071,J5231081,J5231089,J5231103,J5231107,J5231194,J5231221,J5231294,J5231306,

		J5231313,J5231525,J5231625,J8231061,J5231313,J5231525, J5231625,J8231061
--	--	---

2015년 전체환자 표본자료를 이용하여 동일 명세서에 시술코드와 치료재료코드 유무를 교차분석한 결과는 다음과 같다. 스텐트 시술코드가 있는 명세서 208,670건 중 치료재료코드가 없는 경우는 206,978건으로 99.2%나 된다. 이는 급여제한으로 스텐트 시술을 받았지만 치료재료는 청구할 수 없는 경우에 해당된다. 만일 스텐트 시술의 정의를 스텐트 시술코드와 치료재료코드가 모두 있는 경우로 한다면, 급여제한에 따라 스텐트 시술을 받은 환자를 스텐트 시술을 받지 않았다고 오분류할 가능성이 있다. 따라서 청구자료를 이용하여 치료법 또는 결과를 정의할 때는 급여제한으로 인한 오분류의 가능성을 검토해야 한다.

<표 1-13> 스텐트 시술코드와 치료재료코드 교차분석

수가코드		치료재료코드	
		무	유
스텐트 시술 코드	무	135,866,525	181
	유	206,978	1,692

추가적으로 청구자료에서 약물사용을 정의할 경우, 주로 원외처방의 경우 약국명세서 보다 의료기관의 원외처방전 명세서를 이용하는데, 이는 원외처방전을 받은 환자가 반드시 약국에 가서 의약품을 받는다고 가정하는 것이다. 만일 환자가 의료기관에서 받은 원외처방전으로 약국에서 약을 수령하지 않는다면, 약을 복용하지 않은 경우를 약을 복용했다고 오분류하는 경우에 해당한다. 따라서 이런 오분류 가능성을 파악하기 위하여 의료기관 명세서의 원외처방전과 약국명세서의 직접조제 건을 매칭하여 분석을 실시하였다. 진료기간이 2015년인 전체 명세서에서 의료기간의 처방전 내역 테이블에서 원외처방 건수는 총 518,000,426건이었으며, 약국의 처방전내역테이블 건수는 총 503,008,901건으로 의료기간에서 처방된 건 중 약국에서 의약품을 받은 경우는 97.1%로 파악되었다.

여섯째, 청구자료는 급여체계 내에서 수집되는 자료이므로, 분석대상 청구자료의 연도에 따라 의료기술을 포함하여 약물코드, 처치코드 등의 급여기준, 급여횟수 제한 및 변경사항에 대한 확인이 필요하다.

일곱째, 분석에 이용되는 자료기간 내에 수가코드, 약품코드, 치료재료코드의 변경이 있는지 검토가 필요한데, 심사평가원의 홈페이지에 제공되는 수가파일, 약가파일, 치료재료 파일을 확보하여 확인할 수 있다.

여덟째, 사전에 정의한 연구설계 방법 및 통계분석 계획에 따라 청구자료 분석을 수행하기 위해 통계패키지 프로그램에서 프로그래밍을 해야 하는데, 이때 분석자 간

의 프로그램의 상호검토로 오류발생을 교차 검정하는 것이 필요하다.

추가적으로 각 테이블별 주의사항은 다음과 같다⁴⁾.

<표 1-14> 테이블별 주의사항

한글테이블명	영문테이블명	주의사항
명세서일반내역 테이블	T200(TWJHC200)	<ul style="list-style-type: none"> ·청구된 명세서의 일반사항, 환자 및 기관 등의 기본정보가 포함된 테이블 ·기본적인 상병(주상병명과 첫 번째 부상병명만 있음), 내원일수, 요양개시일자 등의 일반적인 현황을 보고자 할 때는 T200만을 사용하기도 함. 상병코드는 한국표준질병사인분류표의 상병분류기호로 입력되어 있음 ·추가 청구를 포함하지 않는 경우, 명세서조인키 변수를 기준으로 동일명세서는 1줄로 되어 있음 ·처방양상 분석 시 퇴원약 처방여부는 ‘요양일수>내원일수’로 알 수 있음. 즉, 요양일수는 수진자를 진료한 총 일수로 퇴원약 등의 투약일수를 포함하나, 내원일수는 수진자가 진료를 받기위해 요양기관에 내원한 일수(초진+재진)또는 입원일수를 말하므로 요양일수가 내원일수보다 큰 경우에는 퇴원약을 처방 받은 것으로 정의함 ·동일 요양기관이라도 건강보험, 의료급여에 따라 요양기관종별이 달라질 수 있음 ·서식구분코드 중 보건기관이라 함은 보건진료소를 제외한 보건기관만 해당함 ·서식구분코드 중 [정신과] 해석 주의 <ul style="list-style-type: none"> - 서식구분코드 중 정신과라 함은 의료급여 정신과 정액을 의미함 - 2008년 정신과 수가제도가 개선됨에 따라, 2009년 1월 진료분부터 의료급여 정신질환 정액수가에 수가코드(09:의료급여 정신과정액 낮병동, 10:의료급여 정신과정액 입원, 11:의료급여 정신과정액 외래)를 부여하고 명세서를 별도 작성하여 청구하도록 개정됨 ·건강보험 혹은 의료급여 환자는 중복으로 보훈 자격을 가질 수 있음. 입원/외래/ 서비스를 이용하더라도 특례 적용 범위 등이 다를 경우 건강보험 혹은 의료급여로 청구되거나 보훈으로 추가 청구되는 경우가 있음 ·청구구분코드(DMD_TP_CD)는 일반청구(0), 보완청구(1), 추가청구(2), 분리청구(3) 등으로 구분됨. 보완청구는 요양기관에서 요양급여비용을 청구하였으나, 심평원에서 심사불능 처리되어 해당 사유를 보완하여 명세서를 재작성하여 다시 청구하는 건임 ·200테이블에는 정상적으로 심사가 완료된 청구건 뿐만아니라 반송된 청구건도 포함되어 있으므로, 중복 청구된 건을 제외하여 분석할 필요가 있음
진료내역 테이블	T300(TWJHC300)	<ul style="list-style-type: none"> ·병원에서 내에서의 처치, 수술, 원내약처방 등의 상세 의료서비스 내역이 포함 된 테이블

4) 황진섭, 신상진, 김종희, 오성희, 강현경, 박승희, 장은진, 최지은, 현민경. 국내 보건의료 이차자료원 활용. 한국보건의료연구원. 2013.

		<p>·처치마다 한 줄로 구성 되어 KEY값 변수를 기준으로 동일명세서에는 여러 줄로 되어있고 줄번호 변수가 있음</p> <p>·상세의료서비스별 비용 산출시, '1_2구분'에서 '가산 있음'일 경우, 요양기관종별 가산율을 고려하여 비용을 산출해야 함</p>
수진자상병내역 테이블	T400(TWJHC400)	<p>·T200의 주상병과 부상병을 포함한 추가 부상병 내역이 포함된 테이블로 T200과 마찬가지로 2011년 1월 1일부터 제 6차 한국표준질병 사인분류체계를 적용. KCD-6차 개정은 후유증 및 합병증 형태 또는 해부학적 위치, 악성 신생물의 진행도에 따라 질병분류가 세분화되어 5단위, 6단위 코드가 증가됨. 또한 206개 상병코드(M54.1: 신경염 또는 신경뿌리염; M08.30~39: 만성 연소성 다발관절염 세분화코드; R84.5: 호흡기관 및 흉부로부터의 검사물의 양성배양 소견; U51-U57: 육경병증 등의 한방 코드 등)는 삭제되었는데 삭제된 코드를 기재하는 경우 심사불능(04) 처리 예정임</p> <p>·제2부상병부터는 중요도 순으로 기재할 것을 「요양급여비용 명세서 작성요령」에서 제시하고 있으나, 요양기관마다 코딩 입력의 차이가 있어 특정 부상병까지만을 고려한 분석시에는(예, 특정 일련번호(SNO)가 2~7까지만 사용) 주의해서 사용할 필요가 있음</p> <p>·주상병 결측 : 통계청 KCD 기준에 없는 경우 결측처리</p>
원외처방전상세 내역 테이블	T530(TWJHC530)	<p>·원외로 처방되는 약에 대한 정보가 있는 테이블로 T300은 원내처방을, T530은 원외 처방전에 나타나는 내용이므로 약물 노출에 대한 분석을 하고자 하는 경우 두 테이블에 있는 정보를 모두 활용하여야 함</p> <p>·처방된 약마다 한 줄로 구성 되어 KEY값 변수를 기준으로 동일명세서에는 여러 줄로 되어 있고 줄번호 변수가 있음. 이때 줄번호는 T300에 있는 줄번호와 다름</p> <p>·T200에서 원외처방일수, 원외처방건수가 0인 아닌 경우라도 원외처방약제비가 0인 경우는 T530 테이블에 해당 명세서 처방전 없음</p> <p>·한 명세서 당 여러 건의 처방전이 포함될 수 있으며, '처방전 교부번호'로 구분할 수 있음</p> <p>·처방전발행일은 처방전교부번호로 산출 할 수 있음. 처방전교부번호의 1-4자리는 발행 된 년도, 5-6자리는 발행월, 7-8자리는 발행일</p> <p>·약물의 경우 2010년부터 통합분류코드(KDI코드)로 변경되었으므로, 분석시 구EDI코드와 KDI코드를 동시에 고려하여야 함</p>

1.6. 자료신청 시 주의사항

자료 신청시 연구과제 수행개요서를 이용하여 해당 연구에 필요한 자료원의 범위에 대해 자료추출 신청을 한다. 이때 연구과제 수행 개요서에 포함되어야 할 내용은 다음과 같다. 세부 항목 신청시 참고할 수 있도록 2015년도 전체 명세서를 대상으로 명세서 현황 분석을 실시하였다. 요양개시년월이 2015년 1월부터 2015년 12월까지이며, 심사년월이 2015년 1월부터 2016년 6월까지인 명세서를 대상으로 지급대상건과 약국을 제외한 921,804,612건의 명세서를 분석한 결과이다.

<표 1-15> 연구과제 수행 개요서 항목

구분	항목	설명
과제정보	연구과제명, 연구기관명, 연구기간, 연구배경 및 목적, 연구내용, 연구방법, 기대효과 및 활용, 선행연구	
요청자료 범위	진료기간	- 2007년부터 신청 가능 - 진료일과 심사일 차이 고려 : 자료추출 시점이 진료기간 이후 3개월 이상이 되도록 신청
	요양기관종별	- 추출 명세서의 요양기관 종별 선택 - 요양기관종별과 서식구분 교차테이블 참고
	대상질환	- 연구대상자 기술 예) 당뇨병 스테로이드 사용 환자
	개인식별정보 포함	- 외부자료 연계가 필요한 경우, 주민등록번호 포함으로 기술
	외부자료 연결	- 외부자료를 반입하여 청구자료와 연계할 경우 기술
	수수료감면	- MOU, 연구중심병원 등 감면 근거 기술
요청자료 상세내역	명세서 일반내역, 진료내역, 수진자 상병내역, 원외처방전 상세내역 등	- 추가로 필요한 테이블 또는 변수가 있을 경우, 기타사항에 표시 - 시군구 지역코드가 필요할 경우 표시
요청자료 산출 조건	진료기간, 요양기관종별	요청자료 범위와 동일하게 표시
	보험자 종별 구분	- 건강보험, 의료급여, 국비보훈 표시 ※ 부록 : 보험자종별 비중 참고
	서식구분	- 요양기관종별 참고하여 표시 ※ 부록 : 요양기관종별과 서식구분의 교차테이블 참고
	요양기관 소재지역	지역을 제한할 경우 기술
	수진자 성별, 연령	성별, 연령 등의 제한이 있을 경우 기술
	상병코드 조건	- 추출조건 : 상병을 기준으로 분석대상 명세서를 추출할 경우 고려할 상병갯수 - 출력조건 : 선택된 명세서의 상병테이블에서 출력할 상병갯수 ※ 부록 : 상병분류별 비중, 명세서 별 상병갯수 분포 참고 - 질환별 상병갯수 정의에 따른 건수 : 전체 상병을 모두 고려했을 경우에 비해 상병갯수 제한에 따른 해당 질환 정의 비율 - 하위코드 포함여부 표시
	약품코드 조건	- 분석대상 명세서 추출시 약품코드를 이용할 경우 해당 약품코드 기술
	행위코드 조건	- 분석대상 명세서 추출시 행위코드 또는 치료재료코드 등 수가코드를 이용할

	경우 해당 코드 기술
서면청구 제외	- 서면청구는 명세서 일반내역 자료만 있으므로, 이를 고려하여 제외여부 결정
DRG청구 제외	- 분석 대상자 및 변수에 DRG로 청구하는 질병군이 있는 경우 제외하면 안 됨
추가청구 제외	- 추가청구 제외 여부 결정 ※ 부록 : 청구구분별 비중 참고
기타 조건	- 분석 대상자에 대한 전체 명세서를 필요한 경우 이를 기술 - 추출 조건을 만족하는 명세서만 필요한 경우 이를 기술 - 동반상병 또는 병용약물 등에 대한 파악이 필요한 경우 : 분석 대상자에 대한 전체 명세서를 신청해야 함

위와 같이 신청자료를 요청할 경우 분석대상자에 대한 정의가 필요하다. 이때 상병코드, 약물코드, 행위 또는 치료재료에 해당하는 수가코드를 이용해서 정의해야 하는데, 이때 다음과 같은 코드 마스터 자료를 이용할 수 있으며, 이는 보건의료개방시스템 홈페이지에서 찾을 수 있다.

<표 1-16> 코드 마스터 테이블

구분	참고 파일
상병코드	상병 마스터파일
약품코드	약가 파일
행위코드	건강보험요양급여비용 파일
치료재료코드	치료재료 목록 파일

산업계에서 자료를 신청할 경우, 자사제품만 분석 가능하며 타사제품의 자료를 신청할 경우에는 자료이용에 대한 동의서를 제출한 후 이용가능하다. 이때 주성분별 의약품코드가 3개 이하인 경우 재식별이 가능하므로, 주성분별 의약품코드가 4개 이상인 경우에만 제공하고 있다.

1.7. 청구자료를 이용한 연구

청구자료를 이용한 연구는 크게 명세서 기반의 분석과 환자 기반의 분석으로 구분할 수 있다. 명세서를 기반으로 한 분석은 주로 질병이나 의료이용에 대한 현황을 파악하기 위해 수행된다. 예를 들어 관심있는 질병으로 인한 연도별 의료이용 현황 및 연도별 특정 의약품의 사용현황 등을 단면적으로 파악하고자 할 때 수행할 수 있다. 세부적으로 명세서를 기반으로 한 분석 방법 및 사례는 2장을 참고하기 바란다.

다음으로 명세서를 이용하여 환자 기준의 분석셋을 구축한 후, 관심 질병의 빈도를 나타내기 위해 발생률, 유병률 등을 산출할 수 있다. 청구자료를 이용하여 발생률, 유병률 등을 산출하는 방법은 3장을 참고하기 바란다.

추가로 청구자료를 이용하여 실제 인구집단에서 이루어지는 일상진료환경에서 노출 또는 치료의 성과를 평가하는 성과연구를 수행할 수 있다. 성과연구 수행시 치료에 따른 성과의 인과성을 설명하기 위해서는 코호트 연구 또는 환자-대조군 연구 등의 연구설계가 필요하며, 관찰연구에서 발생하는 교란요인들을 보정하기 위한 적절한 통계분석방법의 적용이 필요하다. 청구자료를 이용한 성과연구에 대한 방법론적 설명은 4장, 5장을 참고하고, 실제 청구자료를 이용하여 코호트 연구를 수행한 사례는 6장을 참고하기 바란다.

2. 명세서 기반의 분석적 접근

명세서를 기반으로 한 분석은 크게 1) 질병통계 작성, 2) 의약품 사용 분석, 3) 의료경영 관련 지표 생성으로 구분할 수 있다. 2015년 전체환자를 대상으로 하는 NPS-2015 표본자료를 이용하여 명세서 기반의 분석을 위한 단계 및 프로그램을 설명하고자 한다. 2015년 전체환자 데이터셋은 2015년에 요양기관을 방문한 환자를 대상으로 성별, 연령구간(5세 단위)에 따라 층화 계통 추출한 자료로 1,454,252명의 1년간 진료내역과 처방내역을 모두 포함하고 있는 자료이다.

2.1. 명세서 기반의 질병, 의약품 사용 변수 생성

명세서일반내역 테이블(T200)은 하나의 행에 한건의 명세서의 정보를 담고 있으며, 진료내역 테이블(T300), 수진자상병내역 테이블(T400), 원외처방전상세내역 테이블(T530)은 T200 테이블의 한 건의 명세서에 대해 여러 줄로 세부적인 진료내역, 상병, 원외처방내역 등의 정보를 담고 있다.

T200			T300	
spec_id_sno	MSIC_CD	SSICK_CD	SPEC_ID_SNO	DIV_CD
9801	E1150	1109	9801	AA254030
			9801	AA254
			9801	AH200
			9801	AL801
			9801	KK010
			9801	C3710

⇒

T400		T530	
SPEC_ID_SNO	SICK_CD	SPEC_ID_SNO	GNL_NM_CD
9801	M1399	9801	125201AC
9801	E782	9801	268000AT
9801	C1699	9801	438901AT
9801	J459	9801	614100AS
9801	J329		

따라서 명세서 기반의 질병, 의약품 사용에 대한 변수 생성을 위해서는 1) T300, T400, T530 테이블에서 변수를 생성하고, 2) T200 테이블에 조인하기 위해 명세서

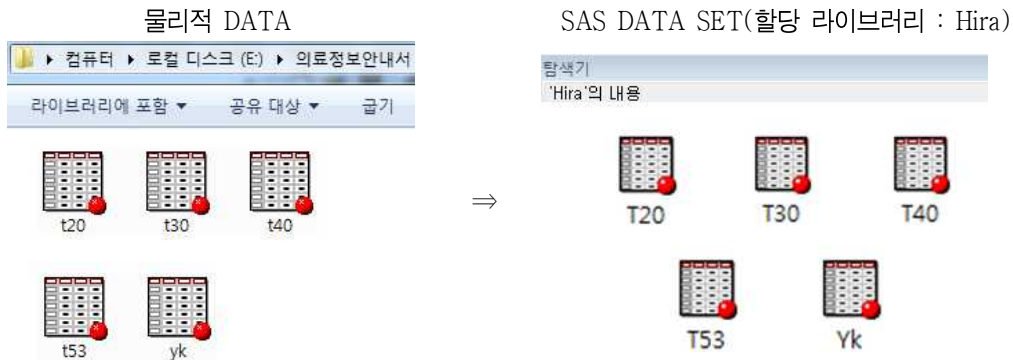
기준 변수를 생성하여 3) T200 테이블에 조인하는 단계를 거쳐야 한다.

※ 프로그램 Tip : 라이브러리 등록

/* SAS 라이브러리명 및 경로 지정 */

libname Hira "E:\의료정보안내서\sasdata"; /* 심평원 2015년 자료 */

위 프로그램 수행 결과 생성되는 DATA SET은 다음과 같다.



① libname HIRA "E:\의료정보안내서\sasdata"

⇒ 위 경로(E:\의료정보안내서)를 HIRA 라이브러리로 할당하여 앞으로 가공할 대상의 DATA SET이 있는 경로를 지정

- libname 라이브러리이름 '경로' : LIBNAME문장은 작업을 할 때마다 영구적 SAS라이브러리에 저장된 SAS파일들에 접근하기 위해 라이브러리를 할당한다.

예)



② options compress=yes

⇒ SAS 데이터 셋 생성 시 compress 옵션을 사용함으로써 데이터 사이즈의 효율적인 관리 및 데이터 작업시간을 단축시킬 수 있다.

- compress = YES : 문자형(CHAR)필드를 압축하는 옵션이다. 데이터의 용량이 낭비되는 현상을 방지할 수 있으며, 공백을 포함한 문자가 2,3byte 연속 반복될 때 유용하다.

□ 질병 변수 생성

T400 테이블에서 다음 질병에 대한 상병코드가 있는지 여부를 파악하고자 한다.

질환	KCD 상병코드
전체 암(Cancer)	C00-C97
당뇨병(Diabetes melitus)	E10-E14
고혈압(Hypertension)	I10-I15
천식(Asthma)	J45-J46
관절증(Arthrosis)	M15-M19
류마티스 관절염(Rheumatoid arthritis)	M05-M06
신부전증(Renal failure)	N17-N19
뇌혈관 질환(Cerebrovascular diseases)	I60-I69
이상지질혈증(Dislipidemia)	E78
허혈성 심장질환(Ischaemic heart diseases)	I20-I25

T200 테이블에는 명세서조인키(SPEC_ID_SNO)와 주상병(MSICK_CD), 제 1 부상병(SSICK_CD)만 저장되어 있으며, T400 테이블에는 한 건의 명세서에 대한 주상병과 모든 부상병이 다음 그림과 같이 여러 줄로 상병(SICK_CD)에 저장되어 있다. 질병의 특성에 따라 질병 정의 시 사용하는 상병의 개수를 제한할 수 있다.

T200

SPEC_ID_SNO	MSICK_CD	SSICK_CD
10039919	C189	K550



주상병과 제1 부상병

T400

SPEC_ID_SNO	SICK_CD
10039919	C189
10039919	K550
10039919	I109
10039919	K650
10039919	N179



주상병과 모든 부상병

(1) T400 테이블에서 질병 여부 변수 생성

T400 테이블에서 질병 여부 변수는 IF 조건문을 사용하여 다음과 같이 생성할 수 있는데, 해당 상병 조건을 만족할 경우 새롭게 생성된 변수에 1을 할당하는 방식으로 생성한다. 이때 IF 조건문에서 상병코드의 개수가 작은 경우 IN 문장을 사용할 수 있으며, 연속적인 상병코드를 가지고 있는 경우 부등식 조건을 이용할 수 있다. 여기서 정의 대상 상병코드가 3단위 코드인지 4단위 코드인지를 구분하여, SUBSTR 함수를 이용하여 조건문에서 비교할 상병코드의 자리수를 정해 준다.

```

/* T400 테이블에서 질병 여부 변수 생성*/
data t40_1;
  set hira.t40;

/*T40의 모든 상병코드에 대해 대상질병 변수 생성*/
  if 'C00' <= substr(SICK_CD,1,3) <= 'C97' then cancer=1; /* 전체 암 */

  if substr(SICK_CD,1,3) in ('E10', 'E11', 'E12', 'E13', 'E14') then diabetes=1; /* 당뇨병 */

  if substr(SICK_CD,1,3) in ('I10', 'I11', 'I12', 'I13', 'I14', 'I15') then hypertension=1;
  /* 고혈압 */

  if substr(SICK_CD,1,3) in ('J45', 'J46') then asthma=1; /* 천식 */

  if substr(SICK_CD,1,3) in ('M15', 'M16', 'M17', 'M18', 'M19')
  then arthrosis=1; /* 관절증 */

  if substr(SICK_CD,1,3) in ('M05', 'M06')
  then rheumatoid_arthritis=1; /* 류마티스 관절염 */

  if substr(SICK_CD,1,3) in ('N17', 'N18', 'N19')
  then renal_failure=1; /* 신부전증 */

  if substr(SICK_CD,1,3) in ('I60', 'I61', 'I62', 'I63', 'I64', 'I65',
  'I66', 'I67', 'I68', 'I69') then Cerebrovascular=1; /* 뇌혈관 질환 */

  if substr(SICK_CD,1,3) in ('E78') then dislipidemia=1; /* 이상지질혈증 */

  if substr(SICK_CD,1,3) in ('I20', 'I21', 'I22', 'I23', 'I24', 'I25')
  then Ischemic=1; /* 허혈성 심장질환 */

run;

```

위 프로그램 수행 결과 생성되는 t40_1 테이블은 다음과 같다.

	SPEC_ID_SNO	SICK_SNO	SICK_DGJSBJT_CD	SICK_CD	DWD_DGJSBJT_CD	cancer	diabetes	hypertension	asthma	arthrosis
29798339	17358194	1 01		R05	01
29798340	17358194	6 01		J459	01	.	.	.	1	.
29798341	17358194	4 01		J42	01
29798342	17358194	5 01		M79108	01
29798343	17358194	3 01		K297	01
29798344	17358194	2 01		J304	01
29798345	17358197	1 05		S9348	05
29798346	17358197	2 05		M7977	05
29798347	17361613	1 01		E785	01
29798348	17361614	4 01		I748	01
29798349	17361614	1 01		E119	01	.	1	.	.	.
29798350	17361614	3 01		E785	01
29798351	17361614	2 01		K769	01
29798352	17361615	1 01		I109	01	.	.	1	.	.
29798353	17362069	2 05		M5456	05
29798354	17362069	3 05		Z5180	05
29798355	17362069	1 01		R521	01

※ 프로그램 Tip

① if 'C00' <= substr(SICK_CD,1,3) <= 'C97' then cancer=1

⇒ T40의 상병코드가 입력되어있는 'SICK_CD' 변수에서 substr함수를 통해 전체 암 상병코드인 C00과 같거나 크고, C97과 같거나 작은 상병코드를 추출해 cancer라는 변수에

1을 넣어 변수를 생성

② if substr(SICK_CD,1,3) in ('E10', 'E11', 'E12', 'E13', 'E14') then diabetes=1

⇒ T40의 상병코드가 입력되어있는 'SICK_CD' 변수에서 substr함수를 통해 해당 질병에 해당하는 코드가 있을 경우 해당 질병 변수에 1이라는 값을 넣어 변수를 생성

③ substr(문자열, 시작위치, 길이)

특정변수에서 시작위치부터 시작으로 길이만큼 문자열을 추출 할 때 사용하는 함수

(예)

SICK_CD		SICK_CD
N051		N05
N001		N00
N19		N19
N178		N17

substr(SICK_CD, 1, 3)
⇒

SICK_CD		SICK_CD
N051		N051
N001		N001
N19		N19
N178		N178

substr(SICK_CD, 1, 4)
⇒

(2) 명세서 기준 질병 변수 생성

(1)에서 생성된 t40_1 테이블은 동일한 명세서 조인키에 대해 여러 줄이 들어 있으므로 T200 테이블에 조인하기 위해서는 하나의 명세서 조인키에 대한 테이블로 생성이 필요하다. 왜냐하면 테이블 간에 조인 시 기준이 되는 변수가 각 테이블에서 중복이 없어야하기 때문이다. 만일 기준 변수의 중복이 있을 경우 테이블 조인 시 중복이 발생하게 된다. (1)에서 생성된 t40_1 테이블을 이용하여 명세서 기준의 질병변수를 생성하기 위해 SQL 문의 GROUP BY 기능을 이용한다.

SPEC_ID_SNO	cancer	diabetes	hypertension	asthma	artrosis
8127040	.	1	1	.	.
8127040	.	1	1	.	.
8127040	.	1	1	.	.
8127040	.	1	1	.	.
8127040	.	1	1	.	.

↓group by(SPEC_ID_SNO)

SPEC_ID_SNO	cancer	diabetes	hypertension	asthma	artrosis
8127040	.	1	1	.	.

```

/*sql문을 통해 명세서단위로 대상질병 변수 요약*/
proc sql;
create table t40_2 as
select SPEC_ID_SNO,
max(cancer) as cancer, /* 전체 암 */
max(diabetes) as diabetes, /* 당뇨병 */
max(hypertension) as hypertension, /* 고혈압 */
max(asthma) as asthma, /* 천식 */
max(arthrosis) as arthrosis, /* 관절증 */
max(rheumatoid_arthritis) as rheumatoid_arthritis, /* 류마티스 관절염 */
max(renal_failure) as renal_failure, /* 신부전증 */
max(Cerebrovascular) as Cerebrovascular, /* 뇌혈관 질환 */
max(dislipidemia) as dislipidemia, /* 이상지질혈증 */
max(Ischaemic) as Ischaemic /* 허혈성 심장질환 */
from t40_1
group by SPEC_ID_SNO; /* 명세서조인키를 기준으로 통합 */
quit;

```

위 프로그램 수행 결과 생성되는 t40_2 테이블은 다음과 같다.

	SPEC_ID_SNO	cancer	diabetes	hypertension	asthma	arthrosis	rheumatoid_arthritis	renal_failure	Cerebrovascular	dislipidemia	Ischaemic
1	1
2	10
3	100	.	1	1	1	.
4	1000
5	10000	.	.	1
6	100000
7	1000000
8	10000000
9	100000001	1	.	.	.
10	100000002	1	.	.	.
11	100000003	1	.	.	.
12	100000004	1	.	.	.
13	100000005	1	.	.	.

※ 프로그램 Tip

① select SPEC_ID_SNO, max(cancer) as cancer, max(diabetes) as diabetes, ...
group by SPEC_ID_SNO

⇒ 동일한 명세서 조인키를 가지는 경우 cancer 변수의 최대값을 cancer 변수에, diabetes 변수의 최대값을 diabetes 변수에 입력

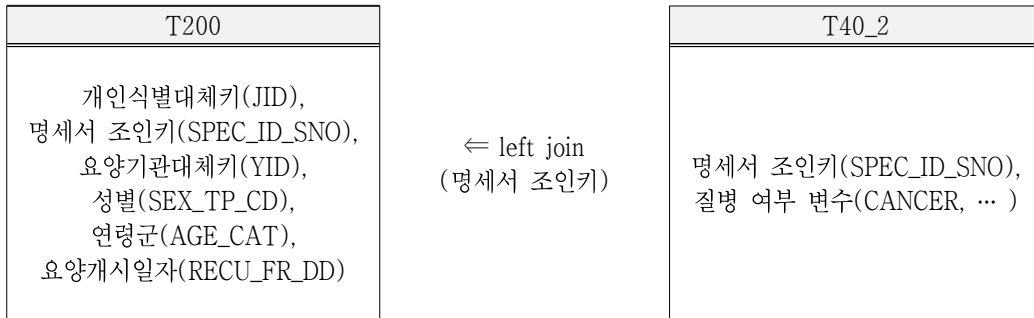
② max(variable) : variable 의 최대값

③ group by variable : variable이 동일한 여러 행들을 모아 한줄로 요약함

(예) select key,
max(var1) as var1_max,
sum(var1) as var1_sum,
min(var1) as var1_min
group by variable;

(3) T200 테이블에 조인

(2)에서 생성된 명세서 기준 질병 여부를 나타내는 변수를 포함한 t40_2 테이블을 명세서 일반내역을 포함하고 있는 T200 테이블에 명세서 조인키 기준으로 조인하여 하나의 테이블을 최종적으로 생성한다. 이때 T200 테이블(왼쪽)을 기준으로 t40_2 테이블을 조인하기 위하여 SQL 문의 LEFT JOIN 기능을 이용한다.



```
/*sql문을 통해 명세서단위로 요약된 대상질병 변수 join*/
proc sql;
  create table t2040 as
  select a.*, b.* /*a: hira.t20 , b: T40_2*/
  from hira.t20 as a
  left join t40_2 as b
  on (a.SPEC_ID_SNO = b.SPEC_ID_SNO);
quit;
```

위 프로그램 수행 결과 생성되는 t2040 테이블은 다음과 같다.

	SPEC_ID_SNO	JID	SEX_TP_CD	PAT_AGE	YID	cancer	diabetes	hypertension	asthma	arthrosis
1	1	1063294	2	52	6150
2	10	460288	1	22	2765
3	100	870968	2	73	9819	.	1	1	.	.
4	1000	1024172	1	74	32523
5	10000	685389	2	80	45080	.	.	1	.	.
6	100000	1370594	2	42	50452
7	1000000	627510	1	2	51434
8	10000000	36014	1	6	34938
9	10000001	1266320	1	57	19294
10	10000002	1266320	1	57	19294
11	10000003	1266320	1	57	19294
12	10000004	1266320	1	57	19294
13	10000005	1266320	1	57	19294

※ 프로그램 Tip

① select a.*, b.*

⇒ a로 지정된 테이블(hira.t20)에 있는 모든 변수와 b로 지정된 질병여부를 포함한 테이블 (t40_2)의 모든 변수를 선택

② from hira.t20 as a

⇒ hira.t20테이블을 a로 지정

③ left join t40_2 as b

⇒ t40_2 테이블을 b로 지정하여 left join

④ on (a.SPEC_ID_SNO = b.SPEC_ID_SNO)

⇒ a로 지정된 테이블(hira.t20) 에서 SPEC_ID_SNO(명세서 조인키)와 b로 지정된 테이블 (t40_2)의 SPEC_ID_SNO(명세서 조인키)를 기준으로 join

⑤ Join : Left join, Right join, Inner join, Full join

a로 지정된 테이블과 b로 지정된 테이블간의 조인키를 기준으로 join
(예)

SELECT <select list>

FROM tableA as A

LEFT/RIGHT/INNER/FULL join tableB as B

on A.KEY = B.KEY

- Left join : 왼쪽테이블을 기준으로 오른쪽 테이블을 결합하는 방법. 왼쪽테이블에 대한 정보를 손실시키지 않고 오른쪽 테이블의 왼쪽테이블과 교차하는 정보만 결합하여 최종테이블을 생성

T200		Left join ⇐	t40_2		
SPEC_ID_SNO	YID		SPEC_ID_SNO	cancer	diabetes
10039919	19294		10039919	.	1
10221567	50452		10615722	1	.
10233106	34938		11757521	1	1
11757521	51434		15348212	1	.

SPEC_ID_SNO	YID	cancer	diabetes
10039919	19294	.	1
10221567	50452	.	.
10233106	34938	.	.
11757521	51434	1	1

- Right join : 오른쪽테이블을 기준으로 왼쪽테이블을 결합하는 방법. 오른쪽테이블에 대한 정보를 손실시키지 않고 왼쪽 테이블의 오른쪽 테이블과 교차하는 정보만 결합하여 최종테이블 생성

T200		Right join ⇒	t40_2		
SPEC_ID_SNO	YID		SPEC_ID_SNO	cancer	diabetes
10039919	19294		10039919	.	1
10221567	50452		10615722	1	.
10233106	34938		11757521	1	1
11757521	51434		15348212	1	.

SPEC_ID_SNO	YID	cancer	diabetes
10039919	19294	.	1
10615722	.	1	.
11757521	.	1	1

15348212	51434	1	.
----------	-------	---	---

- Inner join : 왼쪽테이블과 오른쪽테이블이 교차(KEY값이 둘 다 존재)하는 테이블만 결합하는 방법. 공통 된 값(교집합)만으로 테이블을 생성하게 되기 때문에 두 테이블이 서로 완벽하게 대응되는 KEY값을 모두 갖고 있지 않으면 정보 손실이 발생

T200		Inner join	t40_2		
SPEC_ID_SNO	YID		SPEC_ID_SNO	cancer	diabetes
10039919	19294		10039919	.	1
10221567	50452		10615722	1	.
10233106	34938		11757521	1	1
11757521	51434		15348212	1	.

SPEC_ID_SNO	YID	cancer	diabetes
10039919	19294	.	1
11757521	51434	1	1

- Full join: 왼쪽, 오른쪽 둘 중 하나의 테이블에 KEY값이 존재하면 그 값들을 결합하여 테이블 생성

T200		Full join	t40_2		
SPEC_ID_SNO	YID		SPEC_ID_SNO	cancer	diabetes
10039919	19294		10039919	.	1
10221567	50452		10615722	1	.
10233106	34938		11757521	1	1
11757521	51434		15348212	1	.

SPEC_ID_SNO	YID	cancer	diabetes
10039919	19294	.	1
10615722	.	1	.
10221567	50452	.	.
10233106	34938	.	.
11757521	51434	1	1
15348212	.	1	.

(4) T2040 테이블에 요양기관 정보 조인

추가적으로 요양기관 테이블에 있는 변수들을 T2040 테이블에 조인하여 분석에 활용하고자 할 경우 다음과 같이 요양기관 정보를 조인할 수 있다.

T2040
요양기관대체키(TID), 개인식별대체키(JID), 명세서 조인키(SPEC_ID_SNO), 성별(SEX_TP_CD), 연령군(AGE_CAT), 요양개시일자(RECU_FR_DD), 질병여부변수(CANCER, ...)

⇐ left join
(요양기관 대체키)

YK
요양기관대체키(YID), 요양기관종별(CL_CD), 시도구분(SIDO_CD)

```

/* 요양기관 변수 조인*/
proc sql;
  create table t2040_yk as
  select a.*,
         b.YID, b.SIDO_CD, b.CL_CD
  from t2040 as a
  left join hira.yk as b /*a: t2040, b: hira.yk */
  on (a.YID = b.YID); /* YID(요양기관 대체키)를 기준으로 left join*/
quit;

```

위 프로그램 수행 결과 생성되는 t2040_yk테이블은 다음과 같다.

	SPEC_ID_SNO	YID	SIDO_CD	CL_CD	cancer	diabetes	hypertension
1	10039919	1	11	11	1	1	1
2	10039918	1	11	11	.	.	1
3	10041344	1	11	11	.	1	1
4	10036566	1	11	11	.	.	.
5	10040711	1	11	11	.	1	1
6	10036564	1	11	11	.	.	.
7	10036563	1	11	11	1	.	.
8	10039920	1	11	11	1	1	1
9	10036565	1	11	11	.	.	.
10	10033952	1	11	11	1	.	1
11	10033954	1	11	11	.	1	.

※ 프로그램 Tip

① select a.*, b.YID, b.SIDO_CD, b.CL_CD

⇒ a로 지정된 테이블(t2040)에 있는 모든 변수와 b로 지정된 테이블(hira.yk)의 YID, SIDO_CD, CL_CD 변수를 선택

② from t2040 as a

⇒ t2040 테이블을 a로 지정

③ left join hira.yk as b

⇒ hira 라이브러리의 yk 테이블을 b로 지정

④ on (a.YID = b.YID)

⇒ a로 지정된 테이블(t2040)에서 YID(요양기관 조인키)와 b로 지정된 테이블(hira 라이브러리의 yk)의 YID(요양기관 대체키)를 기준으로 join

□ 의약품 사용 변수 생성

T300 테이블과 T530 테이블에서 다음과 같은 고혈압 약제의 사용량과 사용금액을 파악하고자 한다. 의약품 사용에 대해 파악하기 위해서는 주성분코드 또는 약품코드를 이용할 수 있으며, 약품코드는 T300테이블과 T530테이블의 분류코드(DIV_CD)에 입력되어 있다. 여기서는 주성분코드를 이용하고자 한다.

분류	성분명
ARB	losartan, candesartan, eprosartan, irbesartan, olmesartan, telmisartan, valsartan

T300 테이블에는 요양기관 내에서의 약 처방 외에도 처치, 수술, 청구금액 등 상세 의료서비스 현황이 저장되어있기 때문에 주성분 코드가 입력되어 있는 행을 추출하여 이용해야 한다. T530 테이블에는 원외처방 약제만 저장되어 있다. 따라서 필요에 따라 원내처방과 원외처방을 구분한 후 의약품 테이블을 생성해 의약품 사용에 대한 분석을 할 수 있으며, 표본자료의 특성상 약품코드가 마스킹 되어있고 9단위인 주성분 코드가 8단위임을 고려하여 SUBSTR함수를 이용해 원하는 의약품에 대해 정의할 수 있다.

T300				T530		
SPEC_ID_SNO	DIV_CD	AMT	GNL_NM_CD	SPEC_ID_SNO	AMT	GNL_NM_CD
10000007	AA254011	12190	\$	10000008	2440	111001AT
10000007	AL825	1810	\$	10000008	19520	185701AT
10000007	\$	1035	111402AT			
10000007	\$	4050	459801AT			
10000007	\$	2678	235502AT			

↓

원내 세부 진료 내역

↓

원외처방전상세내역

(1) 의약품 테이블 생성

의약품 테이블은 T300 테이블과 T530테이블을 통합하여 다음과 같이 생성할 수 있는데, WHERE문을 사용하여 T300 테이블에서 주성분코드가 입력되어있는 의약품 처방 내역을 선택하고, 원내·외처방 구분 변수에 1을 할당하여 원내처방을 정의하며, T530 테이블에서 원내·외처방 구분 변수에 2을 할당하여 원외처방을 구분한다. 또한, 사용량 변수명이 다르기 때문에 사용량을 동일한 변수명에 할당하여 생성한다.

T300
명세서 조인키(SPEC_ID_SNO), 주성분코드(GNL_NM_CD), 사용량(TOT_USE_QTY_OR_EXEC_FQ), 금액(AMT)

↑ set

T530
명세서 조인키(SPEC_ID_SNO), 주성분코드(GNL_NM_CD), 사용량(TOT_USE_QTY_EXEC_FQ), 금액(AMT)

⇒

의약품 테이블
명세서 조인키(SPEC_ID_SNO), 주성분코드(GNL_NM_CD), 사용량(TOT_USE), 금액(AMT)

```

/*원내 의약품 사용 변수 생성 - 약물 및 실적(사용량, 금액) 추출 */
data drug_table_30;
set hira.t30;
in_out=1; /*원내외처방 구분변수 1=원내, 2=원외*/
where gnl_nm_cd ne '$'; /*약처방 외의 추가행위를 제외하기 위해 주성분코드가 '$' 인 줄 제거*/
TOT_USE=TOT_USE_QTY_OR_EXEC_FQ; /*T300의 사용량 변수와 T530의 사용량 변수의 변수명
이 다르기 때문에 하나의 변수로 통합*/
keep spec_id_sno gnl_nm_cd amt TOT_USE in_out;
run;

/*원외 의약품 사용 변수 생성 - 약물 및 실적(사용량, 금액) 추출 */
data drug_table_53;
set hira.t53;
in_out=2; /*원내외처방 구분변수 1=원내, 2=원외*/
TOT_USE=TOT_USE_QTY_EXEC_FQ;
keep spec_id_sno gnl_nm_cd amt TOT_USE in_out;
run;

/*의약품 사용 변수 생성 - 의약품 테이블 생성*/
data drug_table;
set drug_table_30 drug_table_53;
run;

```

위 프로그램 수행 결과 생성되는 drug_table 테이블은 다음과 같다.

	SPEC_ID_SNO	AMT	GNL_NM_CD	in_out	TOT_USE
1	9801	30	131801AT	1	15
2	9801	266	242303BI	1	1
3	9803	30	131801AT	1	15
4	10095	266	242303BI	1	1
5	10097	18	131801AT	1	9
6	10098	266	242303BI	1	1
7	13558	12	131801AT	1	6
8	13564	12	131801AT	1	6
9	13568	30	131801AT	1	15
10	14513	313	332700CO	1	1
11	14541	147	131802BI	1	1
12	14542	499	223802BI	1	1
13	14552	499	223802BI	1	1

※ 프로그램 Tip

① in_out=1, in_out=2

⇒ T300의 원내처방과 T530의 원외처방을 통합하기 전 in_out변수에 T300의 경우 1, T530의 경우 2으로 값을 할당하여 원내 · 외 구분

② where gnl_nm_cd ne '\$'

⇒ T300에서 약 처방 외의 수가행위를 제외하기 위하여 주성분코드가 '\$' 인 변수를 제외하고 주성분코드가 입력되어 있는 줄만 선택

- where : 특정 조건에 맞는 데이터 셋을 생성하거나 내보낼 때 사용하는 함수
(예)

gnl_nm_cd		gnl_nm_cd
228713BI	where gnl_nm_cd ne '\$' ⇒	228713BI
142201BI		142201BI
\$		
134103AT		134103AT

③ TOT_USE=TOT_USE_QTY_OR_EXEC_FQ

⇒ T300의 사용량 변수(TOT_USE_QTY_OR_EXEC_FQ)와 T530의 사용량 변수(TOT_USE_QTY_EXEC_FQ)의 변수명이 다르기 때문에 하나의 변수로 통합

(예) 변수명이 다른 경우, 테이블 통합 결과

TOT_USE_QTY_ OR_EXEC_FQ	1	2	1	5
TOT_USE_QTY_ EXEC_FQ	2	1	1	8

set

TOT_USE_QTY_ EXEC_FQ	1	2	1	5
TOT_USE_QTY_ OR_EXEC_FQ
.	2	1	1	9

(예) 변수명이 동일한 경우, 테이블 통합 결과

TOT_USE	1	2	1	5
TOT_USE	2	1	1	8

set

TOT_USE	1	2	1	5	2	1	1	8
---------	---	---	---	---	---	---	---	---

④ keep spec_id_sno gnl_nm_cd amt TOT_USE in_out

⇒ 의약품 사용 분석에 필요한 변수인 명세서 조인키, 주성분코드, 금액, 사용량, 원내·외
처방구분변수만 선택

- keep : 원하는 변수만 설정하여 남김

(예)

SPEC_ID_SNO	in_out	gnl_nm_cd
10000007	1	228713BI
10000008	0	142201BI
10000022	1	\$
10000028	1	134103AT

keep SPEC_ID_SNO

SPEC_ID_SNO	10000007	10000008	10000022	10000028
-------------	----------	----------	----------	----------

- drop : 필요 없는 변수 제거

(예)

SPEC_ID_SNO	in_out	gnl_nm_cd	drop SPEC_ID_SNO ⇒	in_out	gnl_nm_cd
10000007	1	228713BI		1	228713BI
10000008	0	142201BI		0	142201BI
10000022	1	\$		1	\$
10000028	1	134103AT		1	134103AT

(2) 의약품 테이블에서 의약품 변수 생성

(1)에서 생성된 주성분 코드와 사용량, 금액 등을 포함한 drug_table로 고혈압 약제 주성분코드에 해당하는 주성분명 정의는 IF 조건문을 사용하여 다음과 같이 생성할 수 있는데, 해당 주성분코드 조건을 만족할 경우 성분명 변수(gnl_nm)에 성분명을 할당하는 방식으로 생성한다. 이때 표본자료의 특성상 주성분코드는 8단위 코드임을 고려하여 생성한다.

/*의약품 변수 생성 - 고혈압 약제(APB)에 따른 성분명*/

```
data drug_table_1;
  set drug_table;
  length gnl_nm $ 20;
  if gnl_nm_cd in ('185701AT', '185702AT') then gnl_nm='losartan';
  if gnl_nm_cd in ('122601AT', '122602AT') then gnl_nm='candesartan';
  if gnl_nm_cd = '429201AT' then gnl_nm='eprosartan';
  if gnl_nm_cd in ('177301AT', '177303AT') then gnl_nm='irbesartan';
  if gnl_nm_cd in ('468501AT', '468502AT', '468503AT') then gnl_nm='olmesartan';
  if gnl_nm_cd in ('378801AT', '378802AT') then gnl_nm='telmisartan';
  if gnl_nm_cd in ('247101AT', '247102AT', '247103AT', '247104AT') then gnl_nm='valsartan';
run;
```

위 프로그램 수행 결과 생성되는 drug_table_1 테이블은 다음과 같다.

	SPEC_ID_SNO	AMT	GNL_NM_CD	in_out	TOT_USE	gnl_nm
7216	287068	156	249201BI	1	1	
7217	287068	468	249201BI	1	3	
7218	287068	1868	142710BI	1	2	
7219	287069	396	459901AT	1	2	
7220	287069	718	168602BI	1	0,2	
7221	287069	2072	228714BI	1	2	
7222	287069	897000	618801BI	1	1	
7223	287069	22092	222601AT	1	4	
7224	287069	3075	106301AT	1	75	
7225	287069	584	378801AT	1	2	telmisartan
7226	287069	3525	105201AS	1	75	
7227	287069	19200	480401AC	1	50	
7228	287069	240	217001AT	1	16	
7229	287069	360	217001AT	1	24	
7230	287069	240	217001AT	1	16	
7231	287069	360	217001AT	1	24	

※ 프로그램 Tip

① length gnl_nm \$ 20;

⇒ 문자변수의 기본 길이는 8이기 때문에 성분명 변수의 길이를 20으로 지정

- length : 문자변수의 길이 지정

(예)

length문 지정 안 할 경우

gnl_nm
telmisar
losartan
valsarta
valsarta

length gnl_nm \$ 20

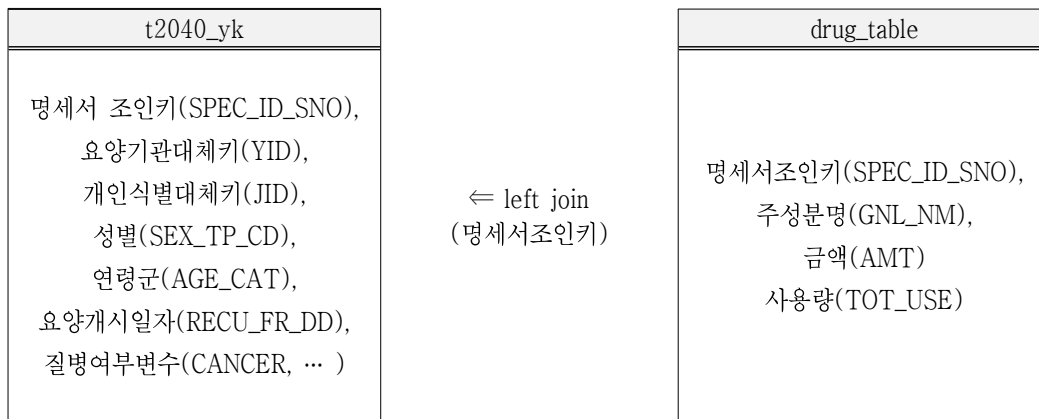
gnl_nm
telmisartan
losartan
valsartan
valsartan

② if gnl_nm_cd in ('185701AT', '185702AT') then gnl_nm='losartan'...

⇒ 의약품 테이블에 주성분코드가 입력되어있는 'gnl_nm' 변수에서 고혈압 약제에 대한 주성분코드에 해당하는 코드가 있을 경우 해당 성분명을 성분명변수에 넣어 변수 생성

□ 질병 및 의약품 사용 통합 테이블 생성

2.1-(4)에서 명세서 기준 요양기관 테이블의 변수를 T200 테이블에 조인하여 생성된 t2040_yk테이블에 주성분 코드와 사용량, 금액 등을 포함한 drug_table 테이블을 조인하여 질병 및 의약품 사용 통합 테이블을 생성한다. 이때 t2040_yk 테이블을 기준으로 drug_table을 조인하기 위하여 SQL문의 LEFT JOIN 기능을 이용한다.



```
proc sql;
  create table t2040_yk_drug as
  select a.* , b.spec_id_sno, b.gnl_nm, b.amt, b.tot_use
  from T2040_yk as a
  left join hira.drug_table as b
  on (a.spec_id_sno = b.spec_id_sno);
quit;
```

위 프로그램 수행 결과 생성되는 t2040_yk_drug 테이블은 다음과 같다.

	SPEC_ID_SNO	JID	SEX_TP_CD	MSICK_CD	SSICK_CD	DGSBJT_CD	RECU_FR_DD	YID	SIDO_CD	CL_CD	AMT	gnl_nm	TOT_USE	cancer	diabetes
1	1	1063294	2	D259	K660	10	20151223	6150	37	11
2	10	460288	1	K352		04	20151221	2765	38	11
3	100	870968	2	E119	I109	01	20151207	9819	21	31
4	1000	1024172	1	H2592		12	20151231	32523	23	31
5	10000	685389	2	I100		01	20150916	45080	34	31
6	100000	1370594	2	N72	N762	15	20151217	50452	11	31
7	1000000	627510	1	J029		01	20151212	51434	26	31
8	10000000	36014	1	J209	J042	23	20150203	34938	11	31
9	10000001	1266320	1	N185	E211	01	20150516	19294	33	31
10	10000002	1266320	1	N185	E211	01	20150518	19294	33	31
11	10000003	1266320	1	N185		01	20150521	19294	33	31
12	10000004	1266320	1	N185	E211	01	20150523	19294	33	31
13	10000005	1266320	1	N185	E211	01	20150525	19294	33	31
14	10000006	1266320	1	N185	B353	01	20150528	19294	33	31
15	10000007	1266320	1	N185	E211	01	20150530	19294	33	31	6240 olmesartan	15	.	.	.
16	10000008	122345	1	I109	I259	01	20150526	19294	33	31	19520 losartan	40	.	.	.

※ 프로그램 Tip

① select a.*, b.spec_id_sno, b.gnl_nm, b.amt, b.tot_use

⇒ a로 지정된 테이블(T2040_yk)에 있는 모든 변수와 b로 지정된 테이블(hira.drug_table)의 명세서조인키, 성분명, 금액, 사용량변수를 선택

② from T2040_yk as a

⇒ T2040_yk 테이블을 a로 지정

③ left join hira.drug_table as b

⇒ hira 라이브러리의 drug_table 테이블을 b로 지정

④ on (a.spec_id_sno = b.spec_id_sno)

⇒ a로 지정된 테이블(T2040_yk)에서 spec_id_sno(개인식별대체키)와 b로 지정된 테이블(hira 라이브러리의 drug_table)의 spec_id_sno(개인식별대체키)를 기준으로 join

2.2. 질병통계 작성

2.1에서 생성한 명세서 기준 질병여부 변수 및 T200 테이블의 명세서의 기본 정보를 이용하여 건강보험통계연보 형식의 진료실 인원, 내원일수, 급여일수, 진료비, 급여비, 환자당 진료비, 환자당 일당 진료비의 합과 평균 등의 기술통계량을 계산하고자 한다. 이를 위해 명세서 기본 정보 중 범주화가 필요한 변수 등에 대해서는 먼저 다음과 같이 변수를 범주화 한다.

(1) 변수 범주화

```
/*질병 DB 만들기 - 최종 분석테이블(연령변수를 범주화 하여 연령군변수 생성)*/
data hira.t20_table;
  set t2040_yk;
/*if문을 통해 연령 범주화(10세 단위)*/
  if 0 <= pat_age <= 9 then age_cat = 0;
  else if 10 <= pat_age <=19 then age_cat = 1;
  else if 20 <= pat_age <=29 then age_cat = 2;
  else if 30 <= pat_age <=39 then age_cat = 3;
  else if 40 <= pat_age <=49 then age_cat = 4;
  else if 50 <= pat_age <=59 then age_cat = 5;
  else if 60 <= pat_age <=69 then age_cat = 6;
```

```

else if 70 <= pat_age <=79 then age_cat = 7;
else if 80 <= pat_age then age_cat = 8;
run;

```

위 프로그램 수행 결과 생성되는 hira.t20_table 테이블은 다음과 같다.

SPEC_ID_SNO	JID	PAT_AGE	age_cat	cancer	diabetes
10039919	410160	65	6	1	1
10039918	640875	67	6	.	.
10041344	1301701	57	5	.	1
10036566	1045901	0.9	0	.	.

※ 프로그램 Tip

① if 0 <= pat_age <= 9 then age_cat = 0

⇒ 연속형변수인 연령을 10세 단위로 범주화 하여 연령군변수 생성, 0세부터 9세까지는 'age_cat' 변수에 0으로 값을 저장

② else if 10 <= pat_age <=19 then age_cat = 1 ...

⇒ if문의 조건에 만족하지 않는 이외의 자료를 탐색하여 10대부터 80대 이상까지 'age_cat' 변수에 값을 저장

- if ① then A else if ② then B : ① 조건을 만족하면 A를 실행하며, ① 조건을 만족하지 않는 경우 else문에서 다시 ②조건을 만족하는지 확인함

(예)

PAT_AGE		if-then else if-then		age_cat
0.9		0 <= pat_age <= 9		0
65		60 <= pat_age <= 69		6
67	→	60 <= pat_age <= 69	→	6
57		50 <= pat_age <= 59		5

(2) 질병통계 작성 1

(1)에서 생성한 hira.t20_table 테이블을 이용하여 연간 특정 질병의 진료실 인원, 내원일수, 요양일수, 진료비, 급여비의 합을 계산하기 위하여 SQL 문을 사용한다. 여기서는 암환자의 입원을 대상으로 한다.

```

/*질병통계(암환자) 작성1*/
proc sql;
create table cancer as
select count(distinct JID) as 진료실인원,
       sum(VST_DDCNT) as 내원일수,
       sum(RECU_DDCNT) as 요양일수,
       sum(RVD_RPE_TAMT_AMT) as 진료비,
       sum(RVD_INSUP_BRDN_AMT) as 급여비
where cancer=1 and FOM_TP_CD='021';/*질병별, 서식구분별*/

```

from hira.t20_table;
quit;
JID:수진자대체키, VST_DDCNT:내원일수, RECU_DDCNT:요양일수, RVD_RPE_TAMT_AMT:심결요양급여비용총액, RVD_INSUP_BRDN_AMT:심결보험자부담금액

위 프로그램 수행 결과 생성되는 cancer 테이블은 다음과 같다.

진료실인원	내원일수	요양일수	진료비	급여비
15,567	488,474	821,239	120,294,830,610	9,385,478,440

※ 프로그램 Tip

① select count(distinct JID) as j_count, sum(VST_DDCNT) as VST_DDCNT,...

⇒ count, distinct 함수를 이용해 j_count(이용자 수) 변수 생성

- distinct(변수명) : 동일한 값을 가지는 2개 이상의 행들에 대해 중복 없이 하나의 행(unique)으로 추출
- count(변수명) : 변수가 가지고 있는 행의 개수를 헤아리는 함수
- count(distinct(변수명)): 변수명이 동일하지 않는 행의 개수를 개산

(예)

JID	1001645	1001645	1267721	1267721
distinct(JID)	⇒	JID	1001645	1267721
		count	(distinct(JID))	⇒
				진료실 인원
				2

② where cancer = 1 and FOM_TP_CD='021'

⇒ hira.T20_table에서 cancer변수가 1이면서 FOM_TP_CD변수가 021(입원)인 행(명세서)을 추출

- where : 특정 조건에 맞는 데이터 셋을 생성하거나 내보낼 때 사용하는 함수

(예)

cancer	1	.	.	1
⇒	where cancer = 1	⇒	cancer	1
				1

(예)

cancer	FOM_TP_CD	⇒ where cancer = 1 and FOM_TP_CD = '021' ⇒	cancer	FOM_TP_CD
1	021		1	021
1	031			
.	021			
1	031			

(3) 질병통계 작성 2

(2)에서 생성한 cancer 테이블을 이용하여 암환자의 입원에 대해 연간 환자당 진료비, 환자당 일당 진료비의 합과 평균 등을 계산하고자 한다. 먼저 암환자 개개인의 연간 내원일수, 진료비, 환자당 일당 진료비를 계산하기 위하여 SQL문을 이용하여 환자당 연간 내원일수, 진료비를 계산하고, 계산된 진료비를 내원일수로 나누어서 환자당 일당진료비를 계산한다.

```
/*질병통계(암환자) 작성 2*/
proc sql;
  create table cancer_2 as
  select  JID,
          sum(VST_DDCNT) as 내원일수,
          sum(RVD_RPE_TAMT_AMT) as 진료비,
          calculated 진료비 / calculated 내원일수 as 환자당 일당 진료비
  from hira.t20_table
  where cancer=1 and FOM_TP_CD='021';/*질병별, 서식구분별*/
  group by jid;
quit;
JID:수진자대체키, VST_DDCNT:내원일수, RVD_RPE_TAMT_AMT:심결요양급여비용총액
```

위 프로그램 수행 결과 생성되는 cancer_2 테이블은 다음과 같다.

JID	내원일수	진료비	환자당 일당 진료비
1000300	8	1,901,820	237,728
1000308	16	2,328,310	145,519
1000443	42	10,437,360	248,509
1000516	62	16,095,550	259,606

위 테이블에서 계산된 환자당 일당 진료비의 기술통계량은 PROC MEANS 프로시저를 이용하여 다음과 같이 계산할 수 있다. 여기서 Q1은 제1사분위수(제25백분위수), Q3은 제3사분위수(제75백분위수)를 나타낸다.

```
/*질병통계(암환자) 작성 3 */
proc means data=cancer_2 n mean sum q1 q3 ;
  var 진료비 환자당 일당 진료비;
```

run;

위 프로그램 수행 결과 생성되는 기술통계량 값은 다음과 같다.

변수	N	mean	sum	Q1	Q3
환자당 진료비	15,567	7,727,554	120,294,830,610	1,606,870	9,466,100
환자당 일당 진료비	15,440	352,310	5,439,660,493	184,552	6,103

※ 프로그램 Tip

- ① select JID, sum(VST_DDCNT) as 내원일수,
sum(RVD_RPE_TAMT_AMT) as 진료비,
calculated 내원일수 / calculated 진료비 as 환자당 일당 진료비
⇒ 암환자의 연간 내원일수와 진료비의 합을 계산하며, 계산된 내원일수를 계산된 진료비로 나누어 환자당 일당 진료비를 계산
- calculated (변수명) : select문에서 기생성한 변수를 나타냄
(예)

내원일수				
2	sum(내원일수) ⇒	내원일수 33	calculated 내원일수 * 2 ⇒	내원일수 66
1				
5				
25				

※ 프로그램 Tip

- ① proc means data=cancer_2 n mean sum q1 q3
⇒ cancer_2 테이블에서 표본 수, 평균, 합계, 제 1사분위수, 제 3사분위수를 계산
- ② var 진료비 환자당 일당 진료비
⇒ 기술통계량을 계산할 변수 지정
- proc means (옵션) : 기술통계량을 구하는 프로시저
(예)

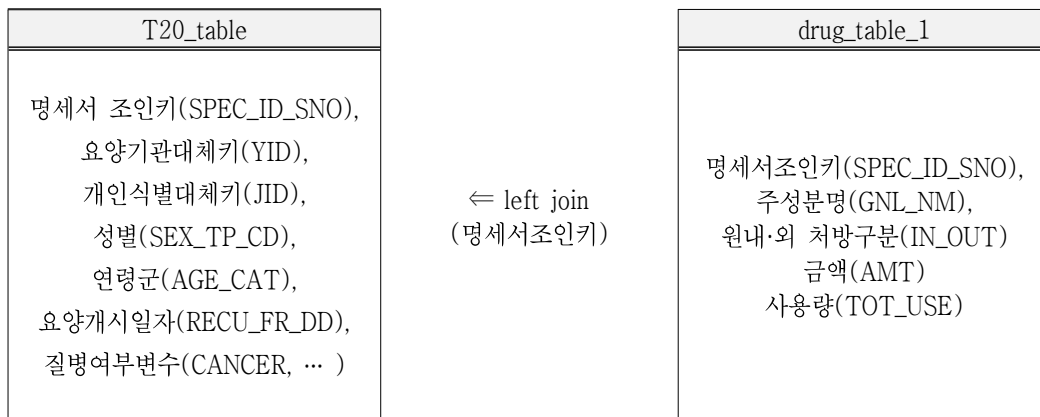
T200				결과			
JID	진료비	급여비	연령	proc means data=cancer_2 n mean sum; var 진료비 급여비 연령; run; ⇒			
1	12,000	10,000	20	변수	N	평균	합계
2	18,000	15,000	67	진료비	4	33013	132050
3	13,150	10,150	55	급여비	4	29013	116050
4	88,900	80,900	42	연령	4	46	51845

2.3. 의약품 사용 분석

다음에서 생성한 drug_table 테이블의 명세서 기준 요양기관 내에서의 의료서비스 현황 및 원외처방 약제 정보를 이용하여 시도별, 요양기관종별, 상병별, 연령별, 원내·외 처방별 주성분별 사용실적 정보를 분석하고자 한다.

□ 고혈압약제 성분 사용 테이블 생성

주성분 코드와 사용량, 금액, 성분명 등을 포함한 의약품 테이블(drug_table_1)과 명세서 기준 질병여부 변수 및 T200 테이블의 명세서를 조인하여 생성한 T20_table 을 조인하기 위하여 SQL문의 LEFT JOIN 기능을 이용한다. 또한 WHERE문을 이용하여 고혈압약제에 대한 성분만을 테이블에 포함시킨다.



```

proc sql;
    create table hira.drug_table as
        select a.*, b.*
        from drug_table_1 as a
        left join hira.T20_table as b
        on (a.spec_id_sno = b.spec_id_sno)
        where a.gnl_nm ne '';
quit;
    
```

위 프로그램 수행 결과 생성되는 hira.drug_table 테이블은 다음과 같다.

	SPEC_ID_SNO	JID	SEX_TP_CD	YID	SIDO_CD	CL_CD	age_cat	RECU_FR_DD	AMT	GNL_NM_CD	in_out	TOT_USE	gnl_nm	cancer	diabetes
1	10000007	1266320	1	19294	33	31	5	20150630	6240	468501AT	1	15	olmesartan	.	.
2	10000008	122345	1	19294	33	31	6	20150626	19520	185701AT	0	40	losartan	.	.
3	10000022	151590	2	19294	33	31	6	20150508	7320	185701AT	1	15	losartan	.	.
4	10000028	151590	2	19294	33	31	6	20150522	7320	185701AT	1	15	losartan	.	1
5	10000037	560727	2	19294	33	31	5	20150511	28260	185702AT	0	30	losartan	.	.
6	10000039	565755	2	19294	33	31	7	20150522	13350	185701AT	0	30	losartan	.	1
7	10000175	482612	2	64518	39	11	7	20150522	401	468501AT	1	1	olmesartan	.	.
8	10000175	482612	2	64518	39	11	7	20150522	2005	468501AT	1	5	olmesartan	.	.
9	10000176	904380	1	64518	39	11	8	20150429	416	468501AT	1	1	olmesartan	.	.
10	10000176	904380	1	64518	39	11	8	20150429	3208	468501AT	1	8	olmesartan	.	.
11	10000178	1381952	2	64518	39	11	7	20150413	2898	177301AT	1	7	irbesartan	1	.
12	10000876	187125	1	63874	21	31	6	20150617	26700	185701AT	0	60	losartan	.	.
13	10000885	1060875	2	63874	21	31	6	20150625	22230	122502AT	0	30	candesartan	.	1
14	10000887	220143	1	63874	21	31	5	20150627	24960	468501AT	0	60	olmesartan	.	1
15	10000890	897868	1	63874	21	31	7	20150608	17190	378802AT	0	30	telmisartan	.	.

※ 프로그램 Tip

① select a.*, b.*

⇒ a로 지정된 테이블(drug_table_1)에 있는 모든 변수와 b로 지정된 테이블(hira.T20_table)의 모든 변수를 선택

② from drug_table_1 as a

⇒ drug_table_1 테이블을 a로 지정

③ left join hira.T20_table as b

⇒ hira 라이브러리의 T20_table 테이블을 b로 지정

④ on (a.spec_id_sno = b.spec_id_sno)

⇒ a로 지정된 테이블(drug_table)에서 spec_id_sno(개인식별대체키)와 b로 지정된 테이블(hira 라이브러리의 T20_table)의 spec_id_sno(개인식별대체키)를 기준으로 join

⑤ where a.gnl_nm ne ''

⇒ 성분명 변수 중 고혈압약제 성분명을 포함하지 않는 행을 제거하여 고혈압약제 성분 사용 테이블 생성

(1) 진료년월, 시도별, 원내·외 처방구분별, 성별, 연령구간별 사용실적

위에서 생성한 hira.drug_table 테이블을 이용하여 진료년월, 시도별, 원내/원외 처방별, 성별, 연령구간별에 따른 고혈압 약제 (ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

```

/* 의약품 사용 분포 분석 1 - 진료년월, 시도별, 원내·외 처방구분별 성별 연령구간별 사용실적*/
proc sql;
create table drug_exam1 as
select substr(recu_fr_dd,1,6) as 요양년월,
       SIDO_CD, CL_CD, in_out, gnl_nm, SEX_TP_CD, age_cat,
       count(distinct jid) as 환자 수
from hira.drug_table
group by recu_fr_ym, SIDO_CD, CL_CD, in_out, gnl_nm, SEX_TP_CD, age_cat;
quit;

```

JID:수진자대체키, RECU_FR_DD:요양개시일자, SIDO_CD:시도코드, CL_CD:요양기관종별코드,
in_out:처방구분, GNL_NM:성분명, SEX_TP_CD:성별, AGE_CAT:연령군

위 프로그램 수행 결과 생성되는 drug_exam1 테이블은 다음과 같다.

	recu_fr_ym	SIDO_CD	CL_CD	in_out	gnl_nm	SEX_TP_CD	age_cat	j_count
1	201501	11	01	0	candesartan	1	1	1
2	201501	11	01	0	candesartan	1	2	2
3	201501	11	01	0	candesartan	1	3	10
4	201501	11	01	0	candesartan	1	4	18
5	201501	11	01	0	candesartan	1	5	29
6	201501	11	01	0	candesartan	1	6	41
7	201501	11	01	0	candesartan	1	7	33
8	201501	11	01	0	candesartan	1	8	5
9	201501	11	01	0	candesartan	2	2	1
10	201501	11	01	0	candesartan	2	3	6
11	201501	11	01	0	candesartan	2	4	5
12	201501	11	01	0	candesartan	2	5	16

※ 프로그램 Tip

- ① select substr(recu_fr_dd,1,6) as 요양년월, SIDO_CD, CL_CD, in_out, gnl_cd, SEX_TP_CD, age_cat, count(distinct JID) as 환자 수
 ⇒ hira.drug_table테이블의 요양년월, 시도코드, 요양기관종별코드, 처방구분, 성분명, 성별, 연령군 변수를 선택
 ⇒ substr함수를 이용하여 recu_fr_dd(요양개시일자)변수에서 요양년월 변수 생성
 ⇒ distinct, count함수를 이용하여 환자 수 변수 생성
- ② group by 요양년월, SIDO_CD, CL_CD, in_out, gnl_nm, SEX_TP_CD, age_cat
 ⇒ 요양월, 시도코드, 요양기관종별코드, 처방구분, 성분명, 성별 변수를 기준으로 데이터를 요약

(2) 시도별 요양기관종별 사용실적

위에서 생성한 hira.drug_table 테이블을 이용하여 시도별 요양기관종별 고혈압 약제 (ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

/* 의약품 사용 분포 분석 2 - 시도별 요양기관그룹별 사용실적*/					
proc sql;					
create table drug_exam2 as					
select SIDO_CD, CL_CD, gnl_nm,					
sum(TOT_USE) as 사용량,					
sum(amt) as 금액,					
count(distinct JID) as 환자 수					
from hira.drug_table					
group by SIDO_CD, CL_CD, gnl_nm;					
quit;					
JID:수진자대체키, SIDO_CD:시도코드, CL_CD:요양기관종별코드, GNL_NM:성분명, TOT_USE:사용량, AMT:금액					

위 프로그램 수행 결과 생성되는 drug_exam2 테이블은 다음과 같다.

시도코드	요양기관종별	성분명	사용량	금액	환자 수
11	1	candesartan	265,679	161,305,578	1,087
11	1	eprosartan	37,149	29,147,574	148
41	72	olmesartan	420	174,720	2

41	72	valsartan	780	354390	3
----	----	-----------	-----	--------	---

※ 프로그램 Tip

- ① select SIDO_CD, CL_CD, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수
⇒ count, distinct함수를 이용해 환자 수 변수 생성
⇒ sum 함수를 이용해 사용량, 금액 변수 생성
② group by SIDO_CD, CL_CD, gnl_nm
⇒ 시도코드, 요양기관종별코드, 성분명 변수를 기준으로 데이터를 요약

(3) 요양기관종별 사용실적

위에서 생성한 hira.drug_table 테이블을 이용하여 요양기관종별 고혈압 약제(ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

```
/* 의약품 사용 분포 분석 3 - 요양기관종별 사용실적*/
proc sql;
create table drug_exam3 as
select CL_CD, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수
from hira.drug_table
group by CL_CD, gnl_nm;
quit;
```

JID:수진자대체키, CL_CD:요양기관종별코드, GNL_NM:성분명, AGE_CAT:연령군, TOT_USE:사용량, AMT:금액

위 프로그램 수행 결과 생성되는 drug_exam3 테이블은 다음과 같다.

요양기관종별	성분명	사용량	금액	환자 수
1	candesartan	594,125	350,402,546	2,595
1	eprosartan	78,010	61,078,675	320
92	telmisartan	33	13,818	2
92	valsartan	302	158,550	2

※ 프로그램 Tip

- ① select CL_CD, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수
⇒ count, distinct함수를 이용해 환자 수 변수 생성
⇒ sum 함수를 이용해 사용량, 금액 변수 생성
② group by CL_CD, gnl_nm
⇒ 요양기관종별코드, 성분명 변수를 기준으로 데이터를 요약

(4) 요양기관종별 원내·외 처방구분별 사용실적

위에서 생성한 hira.drug_table 테이블을 이용하여 요양기관종별 원내·외 처방별 고혈압 약제(ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

/* 의약품 사용 분포 분석 4 - 요양기관종별 원내·원외 사용실적*/					
proc sql;					
create table drug_exam4 as					
select CL_CD, in_out, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,					
count(distinct JID) as 환자 수					
from hira.drug_table					
group by CL_CD, in_out, gnl_nm;					
quit;					
JID:수진자대체키, CL_CD:요양기관종별코드, in_out:처방구분, GNL_NM:성분명, TOT_USE:사용량,					
AMT:금액					

위 프로그램 수행 결과 생성되는 drug_exam4 테이블은 다음과 같다.

요양기관종별	처방구분	성분명	사용량	금액	환자 수
1	0	candesartan	559,293	336,218,191	2,246
1	0	eprosartan	75,292	59,103,855	296
92	1	candesartan	22	10,252	1
92	1	telmisartan	3	1,278	1

※ 프로그램 Tip

① select CL_CD, in_out, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수

⇒ count, distinct함수를 이용해 환자 수 변수 생성

⇒ sum 함수를 이용해 사용량, 금액 변수 생성

② group by CL_CD, in_out, gnl_nm

⇒ 요양기관종별코드, 처방구분, 성분명 변수를 기준으로 데이터를 요약

(5) 성분별 시도별 사용실적

위에서 생성한 hira.drug_table 테이블을 이용하여 성분별 시도별 고혈압 약제(ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

/* 의약품 사용 분포 분석 5 - 성분별 시도별 사용실적*/					
proc sql;					
create table drug_exam5 as					
select gnl_nm, SIDO_CD, sum(TOT_USE) as 사용량, sum(amt) as 금액,					
count(distinct JID) as 환자 수					
from hira.drug_table					
group by gnl_nm, SIDO_CD;					
quit;					
JID:수진자대체키, SIDO_CD:시도코드, GNL_NM:성분명, TOT_USE:사용량, AMT:금액					

위 프로그램 수행 결과 생성되는 drug_exam5테이블은 다음과 같다.

성분명	시도코드	사용량	금액	환자 수
candesartan	11	643,979	377,054,006	2,675
candesartan	21	229,494	128,510,138	1,047
valsartan	39	22,624	15,142,831	129
valsartan	41	3,249	1,818,967	24

※ 프로그램 Tip

① select gnl_nm, SIDO_CD, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수

⇒ count, distinct함수를 이용해 환자 수 변수 생성

⇒ sum 함수를 이용해 사용량, 금액 변수 생성

② group by gnl_nm, SIDO_CD

⇒ 성분명, 시도코드 변수를 기준으로 데이터를 요약

(6) 원내·외 처방구분별 환자 수 사용실적

위에서 생성한 hira.drug_table 테이블을 이용하여 원내·외 처방별 환자 수 및 고혈압 약제(ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

/* 의약품 사용 분포 분석 6 - 원내·원외 환자수 사용실적*/				
proc sql;				
create table drug_exam6 as				
select in_out, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,				
count(distinct JID) as 환자 수				
from hira.drug_table				
group by in_out, gnl_nm;				
quit;				
JID:수진자대체키, in_out:처방구분, GNL_NM:성분명, TOT_USE:사용량, AMT:금액				

위 프로그램 수행 결과 생성되는 drug_exam6테이블은 다음과 같다.

처방구분	성분명	사용량	금액	환자 수
0	candesartan	2298847	1,350,907,590	9,643
0	eprosartan	199695	156,760,229	756
1	telmisartan	178443	52,236,775	1,748
1	valsartan	230798	130756085	2481

※ 프로그램 Tip

① select in_out, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수

⇒ count, distinct함수를 이용해 환자 수 변수 생성

⇒ sum 함수를 이용해 사용량, 금액 변수 생성

② group by in_out, gnl_nm

⇒ 처방구분, 성분명 변수를 기준으로 데이터를 요약

(7) 상병별 연령구간별 사용실적

2.1에서 생성한 t2040_yk_drug 테이블을 이용하여 암과 당뇨병 환자를 대상으로 의약품성분의 연령구간별 고혈압 약제(ARB) 사용실적을 구하기 위하여 SQL문을 이용한다.

```
/*의약품 사용 분포 분석7 - 상병별 연령구간별 사용실적 */
proc sql; /* 전체 암에 대한 상병별 연령구간별 사용실적 */
create table drug_exam7 as
select age_cat, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
       count(distinct JID) as 환자 수
from t2040_yk_drug
where gnl_nm ne " " and cancer=1
group by age_cat, gnl_nm;
quit;
```

```
proc sql; /* 당뇨병에 대한 상병별 연령구간별 사용실적 */
create table drug_exam7 as
select age_cat, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
       count(distinct JID) as 환자 수
from t2040_yk_drug
where gnl_nm ne " " and diabetes=1
group by age_cat, gnl_nm;
quit;
```

JID:수진자대체키, GNL_NM:성분명, AGE_CAT:연령군, TOT_USE:사용량, AMT:금액

위 프로그램을 살펴 보면, 분석대상 질병만 암과 당뇨병으로 차이가 있을을 알 수 있다. 따라서 일부 변수값만 변경이 되고 동일한 프로그램이 반복되는 경우는 아래와 같이 매크로(MACRO)를 이용하여 프로그램을 작성할 수 있다.

```
/*의약품 사용 분포 분석7 - 상병별 연령구간별 사용실적(MACRO)*/
%macro drug_exam7(sick=); /*매크로 및 매크로 변수 생성 */

proc sql;
create table drug_exam7_&sick. as
select age_cat, gnl_nm, sum(tot_use) as tot_use, sum(amt) as amt,
       count(distinct jid) as j_count
from hira.drug_table
where &sick. = 1
group by age_cat, gnl_nm;
quit;

%mend;

options mlogic mprint;
```

```
%drug_exam7(sick=cancer) /* 암에 대한 연령구간별 사용실적 매크로 실행 */
%drug_exam7(sick=diabetes) /* 당뇨병에 대한 연령구간별 사용실적 매크로 실행 */
```

위 프로그램 수행 결과 생성되는 drug_exam7테이블은 다음과 같다.

상병명	연령군	성분명	사용량	금액	환자수
cancer	1	irbesartan	61	58,260	1
cancer	2	irbesartan	112	99,064	2
diabetes	8	telmisartan	63,578	28,792,792	348
diabetes	8	valsartan	96,812	57,029,261	562

※ 프로그램 Tip

- ① select disease, age_cat, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수
⇒ count, distinct함수를 이용해 환자 수 변수 생성
⇒ sum 함수를 이용해 사용량, 금액 변수 생성
- ② where gnl_nm ne " and cancer=1
where gnl_nm ne " and diabetes=1
⇒ t2040_yk_drug 테이블에서 암과 당뇨를 질병으로 앓고 있는 줄 선택 및 고혈압약제를 사용하지 않은 줄을 제거
- ③ group by age_cat, gnl_nm
⇒ 연령군, 성분명 변수를 기준으로 데이터를 요약

※ 프로그램 Tip(macro)

- ① %macro drug_exam6(sick=)
⇒ 매개변수를 가진 매크로 정의, 괄호 내에 정의된 매크로 변수는 매크로 내의 정보를 전달하도록 한다.
(예) 모듈단위로 프로그램을 작성하고, 실행
%MACRO 문장이름(변수1=, 변수2=, ... 변수n=);
SAS 프로그램 및 매크로 프로그램
%MEND 문장이름;
%문장이름(변수1=, 변수2=, ... 변수n=)
- ② proc sql; create table drug_exam6_&sick.as ...
⇒ 매크로가 수행될 때, 매크로 변수는 매크로 내의 SAS프로그램을 호출된 매크로의 지정된 값에 대응시킨다.
- ③ %drug_exam6(sick=cancer)
⇒ 다음과 같이 매크로를 호출하여 매크로 내의 SAS프로그램에 매크로 변수 값을 대응해 SAS 프로그램이 수행하도록 한다.
- ④ options mlogic mprint;
⇒ macro 디버깅 옵션

- MLOSIC | NOMLOSIC : 매크로 부분에 대한 실행 내용을 로그 창에 자세히 출력
- MPRINT | NOMPRINT : 매크로 프로세서에 의해 번역된 SAS코드 출력
- SYMBOLGEN | NOSYMBOLGEN : 매크로 변수의 값을 로그 창에 출력

(8) 상병별 요양기관종별 사용실적

2.1에서 생성한 t2040_yk_drug테이블에 대해 암과 당뇨병환자를 대상으로 고혈압 약제(ARB)의 요양기관종별 사용실적을 구하기 위하여 SQL문을 이용한다.

```
/* 의약품 사용 분포 분석8 - 상병별 요양기관종별 사용실적 */
proc sql; /* 전체 암에 대한 상병별 요양기관종별 사용실적 */
create table drug_exam8 as
select CL_CD, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
       count(distinct JID) as 환자 수
from t2040_yk_drug
where gnl_nm ne "" and cancer=1
group by CL_CD, gnl_nm;
quit;

proc sql; /* 당뇨병에 대한 상병별 요양기관종별 사용실적 */
create table drug_exam8 as
select CL_CD, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
       count(distinct JID) as 환자 수
from t2040_yk_drug
where gnl_nm ne "" and diabetes=1
group by CL_CD, gnl_nm;
quit;
```

JID:수진자대체키, CL_CD:요양기관종별코드, GNL_NM:성분명, TOT_USE:사용량, AMT:금액

위 프로그램 수행 결과 생성되는 drug_exam8테이블은 다음과 같다.

상병명	요양기관종별	성분명	사용량	금액	환자 수
cancer	01	candesartan	12,465	6,664,795	148
cancer	01	eprosartan	804	591,666	8
diabetes	92	candesartan	757	368,932	3
diabetes	92	telmisartan	3	1,278	1

※ 프로그램 Tip

① select CL_CD, gnl_nm, sum(TOT_USE) as 사용량, sum(amt) as 금액,
count(distinct JID) as 환자 수

⇒ count, distinct함수를 이용해 환자 수 변수 생성

⇒ sum 함수를 이용해 사용량, 금액 변수 생성

② where gnl_nm ne "" and cancer=1

where gnl_nm ne "" and diabetes=1

⇒ t2040_yk_drug 테이블에서 암과 당뇨를 질병으로 앓고 있는 줄 선택 및 고혈압약제를 사용

하지 않은 줄을 제거

③ group by CL_CD, gnl_nm

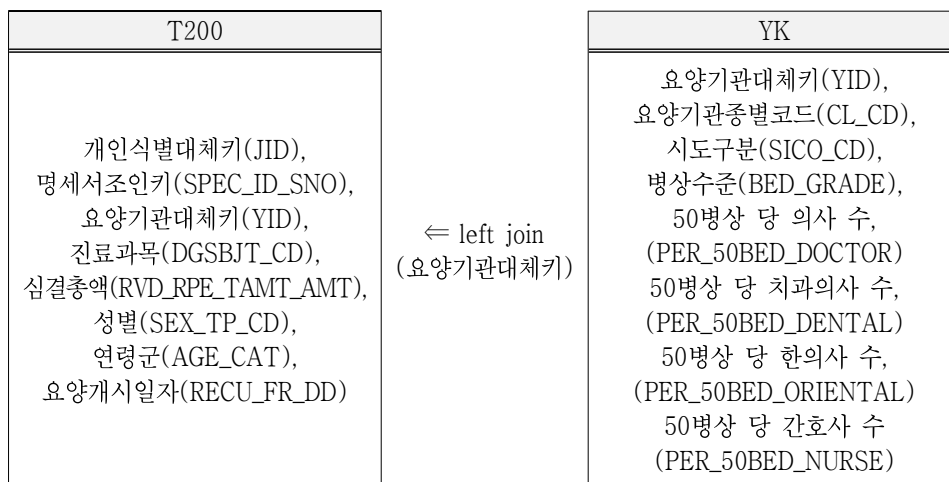
⇒ 요양기관종별코드, 성분명 변수를 기준으로 데이터를 요약

2.4. 의료경영지원 관련 지표 생성

의료경영지원 관련 지표를 생성하기 위해 요양기관 테이블의 요양기관종별코드, 시도구분, 병상수준 등의 정보를 사용하고자 한다. 요양기관의 명세서 일반내역을 포함하고 있는 T200 테이블(왼쪽)을 기준으로 요양기관 테이블을 조인하기 위하여 SQL 문의 LEFT JOIN 기능을 이용한다.

(1) T200 테이블에 요양기관 정보 조인

요양기관 테이블에 있는 변수들을 T200 테이블에 조인하여 분석에 활용하고자 할 경우 다음과 같이 요양기관 정보를 조인할 수 있다.



```
/*의료경영지원 지표 - 요양기관 정보 및 지표 join*/
proc sql;
  create table hira.yk_table as
  select a.JID, a.SPEC_ID_SNO, a.YID, a.DGSBJT_CD, a.RVD_RPE_TAMT_AMT,
        a.RVD_INSUP_BRDN_AMT, a.SEX_TP_CD, a.age_cat, a.RECU_FR_DD,
        b.yid, b.CL_CD, b.SICO_CD, b.BED_GRADE, b.PER_50BED_DOCTOR,
        b.PER_50BED_DENTAL, b.PER_50BED_ORIENTAL, b.PER_50BED_NURSE
  from hira.t20_table as a
  left join hira.yk as b
  on(a.YID = b.YID);
quit;
```

위 프로그램 수행 결과 생성되는 yk_table 테이블은 다음과 같다.

	JID	SPEC_ID_SNO	YID	CL_CD	SIDO_CD	BED_GRADE	PER_50BED_DOCTOR	PER_50BED_NURSE	
1	476282	20587963	1	11	11	14	23		44
2	869650	20595691	1	11	11	14	23		44
3	275199	20724581	1	11	11	14	23		44
4	682971	3574480	1	11	11	14	23		44
5	67061	20597892	1	11	11	14	23		44
6	489636	2076170	1	11	11	14	23		44
7	1369566	20597259	1	11	11	14	23		44
8	481357	2162629	1	11	11	14	23		44
9	489711	2032755	1	11	11	14	23		44
10	122864	20723531	1	11	11	14	23		44
11	1349201	3426350	1	11	11	14	23		44
12	696890	20712397	1	11	11	14	23		44
13	650726	2089226	1	11	11	14	23		44

※ 프로그램 Tip

- ① select a.JID, a.SPEC_ID_SNO, a.YID, a.DGSBJT_CD, a.RVD_RPE_TAMT_AMT,
a.RVD_INSUP_BRDN_AMT, a.SEX_TP_CD, a.age_cat, a.RECU_FR_DD ...
⇒ a로 지정된 테이블(hira.t20_table)에 있는 변수와 b로 지정된 요양기관테이블(YK)의 변수를 선택
- ② from hira.t20 as a
⇒ hira.t20테이블을 a로 지정
- ③ left join hira.yk as b
⇒ hira.yk 테이블을 b로 지정하여 left join
- ④ on (a.YID = b.YID)
⇒ a로 지정된 테이블(hira.t20)에서 YID(요양기관대체키)와 b로 지정된 테이블(hira.t20)의 YID(요양기관대체키)를 기준으로 left join

(2) 지역 내 개원가수 분포

(1)에서 생성한 yk_table 테이블을 이용하여 진료년월별, 시도구분별, 요양기관종별 요양기관 수를 구하기 위하여 SQL 문을 사용한다.

/*의료경영지원 지표 - 지역 내 개원가수*/

```
proc sql;
create table yk_exam1 as
select substr(recu_fr_dd,1,6) as 진료년월,
       SIDO_CD,
       CL_CD, count(distinct YID) as 요양기관 수
from hira.yk_table
group by 진료년월, SIDO_CD, CL_CD;
quit;
```

REC_U_FR_DD:요양개시일자, SIDO_CD:시도코드, CL_CD:요양기관종별코드, YID:요양기관대체키

위 프로그램 수행 결과 생성되는 yk_exam1 테이블은 다음과 같다.

진료년월	시도코드	요양기관종별코드	요양기관 수
201501	11	01	14
201501	11	11	42
201512	41	73	2

201512	41	93	22
--------	----	----	----

※ 프로그램 Tip

- ① select substr(recu_fr_dd,1,6) as 진료년월, SIDO_CD, CL_CD, count(distinct YID) as ...
 ⇒ yk_table 테이블의 진료년월, 시도코드, 요양기관종별 변수를 선택
 ⇒ count, distinct 함수를 이용해 요양기관 수 변수 생성
- ② group by 요양년월, SIDO_CD, CL_CD
 ⇒ 진료년월, 시도코드, 요양기관종별코드를 기준으로 요양기관 수 요약

(3) 일평균환자 수 분포

(1)에서 생성한 yk_table 테이블을 이용하여 시도별, 요양기관종별, 요양기관별, 하루 환자수의 평균을 구하기 위하여 SQL 문을 사용한다.

/*의료경영지원 지표 - 일평균 환자수*/

/* 시도별, 요양기관종별, 요양기관별, 일 환자수 계산*/

proc sql;

create table yk_exam2 as

select SIDO_CD,
 CL_CD,
 YID,
 recu_fr_dd,
 count(distinct JID) as 환자 수

from hira.yk_table

group by SIDO_CD, CL_CD, YID, recu_fr_dd;

quit;

RECU_FR_DD:요양개시일자, SIDO_CD:시도코드, CL_CD:요양기관종별코드, YID:요양기관대체키, JID:수진자대체키

위 프로그램 수행 결과 생성되는 yk_exam2 테이블은 다음과 같다.

시도코드	요양기관종별코드	요양기관 조인키	요양개시일자	환자수
11	01	15012	20150101	21
11	01	15012	20150102	81
41	93	95	20151222	1
41	93	95	20151229	1

※ 프로그램 Tip

- ① select SIDO_CD, CL_CD, YID, recu_fr_dd, count(distinct JID) as 환자 수
 ⇒ yk_table 테이블의 시도코드, 요양기관종별코드, 요양기관 조인키, 요양개시일자를 선택
 ⇒ count, distinct 함수를 이용해 환자 수 변수 생성
- ② group by recu_fr_dd, YID, SIDO_CD, CL_CD
 ⇒ 시도코드, 요양기관종별코드, 요양기관 조인키, 요양개시일자를 기준으로 환자 수 요약

위에서 생성한 환자 수 변수를 이용하여 시도별, 요양기관종별 일평균환자 수를 계산한다.

```
/*의료경영지원 지표 - 일평균 환자 수*/
/* 시도별, 요양기관종별 일평균 환자 수 계산*/
proc means;
class SIDO_CD CL_CD; /* 시도별, 요양기관종별 조합에 대한 기술통계량 산출*/
var j_count; /* 위의 yk_exam2 테이블의 환자 수에 하여 기술통계량 산출*/
run;
SIDO_CD:시도코드, CL_CD:요양기관종별코드
```

위 프로그램 수행결과 생성되는 결과 값은 다음과 같다.

시도코드	요양기관종별코드	일평균환자 수
11	01	103
11	11	23
41	73	1
41	93	2

※프로그램 Tip

① proc means

class SIDO_CD CL_CD

var j_count

⇒ 기술통계량을 구하는 SAS프로시저 중 하나이며, 기술통계량뿐만 아니라 데이터 오류 확인 시에도 많이 사용됨

CLASS에서 지정한 변수들의 수준별 모든 조합에 대해 기술통계량 산출

예) PROC MEANS option;

VAR/BY/CLASS/FREQ/WEIGHT/ID variable;

(4) 시도별 성별 연령군별 환자 수 분포

(1)에서 생성한 yk_table 테이블을 이용하여 시도별, 성별, 연령군별 환자 수를 구하기 위하여 SQL 문을 사용한다.

```
/*의료경영지원 지표 - 시도별 성별 연령군별 환자 수*/
proc sql;
create table yk_exam4 as
select SIDO_CD,
SEX_TP_CD,
age_cat, count(distinct JID) as 환자 수
from hira.yk_table
group by SIDO_CD, SEX_TP_CD, age_cat;
quit;
```

SEX_TP_CD:성별, AGE_CAT:연령군, SIDO_CD:시도코드, JID:수진자대체키

위 프로그램 수행 결과 생성되는 yk_exam4 테이블은 다음과 같다.

시도코드	성별	연령군별	환자 수
11	1	0	17,537
11	1	1	18,887
41	2	7	272
41	2	8	133

※프로그램Tip

- ① select SIDO_CD, SEX_TP_CD, age_cat, count(distinct JID) as 환자 수
 ⇒ hira.yk_table 테이블의 시도코드, 성별, 연령군 선택
 ⇒ count, distinct 함수를 이용해 환자 수 생성
- ② group by SIDO_CD, SEX_TP_CD, age_cat
 ⇒ 시도코드, 성별, 연령군을 기준으로 각 변수들의 값에 따라 환자 수 계산

(6) 시도별 진료과목별 환자 수 분포

(1)에서 생성한 yk_table 테이블을 이용하여 시도별 진료과목별 환자 수를 구하기 위하여 SQL 문을 사용한다.

```

/*의료경영지원 지표 - 시도별 진료과목별 환자 수*/
proc sql;
  create table yk_exam5 as
  select SIDO_CD, dgsbjt_cd, count(distinct JID) as 환자 수
  from hira.yk_table
  where dgsbjt_cd ne '$'
  group by SIDO_CD, dgsbjt_cd;
quit;
SIDO_CD:시도코드, DGSBJT_CD:진료과목코드, CL_CD:요양기관종별코드, YID:요양기관대체키, JID:
수진자대체키

```

위 프로그램 수행 결과 생성되는 yk_exam5 테이블은 다음과 같다.

시도코드	진료과목코드	환자 수
11	00	9,020
11	01	249,985
41	85	803
41	86	4

※프로그램 Tip

- ① select SIDO_CD, DGSBJT_CD, count(distinct JID) as 환자수

⇒ hira.yk_table테이블의 시도코드, 진료과목코드 변수를 선택

⇒ count, distinct 함수를 이용해 환자 수 생성

② **where dgsbjt_cd ne '\$'**

⇒ 진료과목오류코드(\$) 제외

③ **group by SIDO_CD, DGSBJT_CD**

⇒ 시도코드, 진료과목코드를 기준으로 각 변수들의 값에 따라 환자 수 계산

3. 환자 기반의 분석적 접근

3.1. 대푯값, 분율, 비율

연구자료 분석의 첫 단계는 각 측정 결과나 변수들의 분포를 기술하는 것이다. 대푯값은 자료를 대표하여 하나로 표현한 값이며, 평균, 중앙값, 최빈값 등이 있다. 평균(mean) 중에서 가장 많이 사용하는 산술평균은 총 관찰값의 합을 전체 대상수(n)로 나누어 산출한다. 중앙값(median)은 n개의 대상수가 있을 때 크기순으로 배열할 경우 중앙에 있는 값이다. 예를 들어 관찰값이 9개일 경우, 중앙값은 5번째 순위의 값이고, 관찰값이 10개일 때는 5, 6번째 값의 평균이 중앙값이 된다. 최빈값(mode)은 관찰값 중 도수가 가장 많은 값이 된다. 평균과 중앙값은 자료의 분포 모양에 대한 정보를 제공한다. 정규분포를 따르는 경우 평균과 중앙값이 동일하지만, 왼쪽으로 기울어진 자료의 경우 중앙값이 평균보다 작게 된다.

인구집단의 특성을 요약할 때, 비(ratio), 분율(proportion), 비율(rate) 등을 이용하여 제시하게 된다.

비(ratio) 중, 단순비(simple ratio)는 두 사건의 상대적인 크기를 비교한 것으로 A에 대한 B의 비를 B/A로 계산하며, 0보다 큰 값으로 표현한다. 대표적으로 성비 (남자 수/여자 수), 1인당 지방 섭취량 (총 지방 소비량/인구수), 체질량지수(BMI, 체중(kg)/신장(cm)²) 등이 있다.

대응비(Odds)는 동일한 조건하에 있는 집단에서 특정 현상이 있는 사람의 수와 없는 사람의 수에 대한 단순비이다. 예를 들어 N부서의 직원 9명 중 흡연자가 6명 비흡연자가 3명인 경우 흡연자의 대응비는 2로 '이 집단에서 흡연자는 비흡연자 1명에 대해 2명 꼴로 있다'라고 표현할 수 있다.

분율(proportion)은 비(ratio)의 한 형태로 분자가 반드시 분모에 포함되는 경우이다. 전체에서 특정 사건을 경험한 개체가 차지하는 정도를 의미하고, 0과 1사이 값을 가진다. 예를 들어 전체 입원환자 100명 중 암으로 입원한 환자가 10명일 때 암 입원 분율은 $10/100 = 10\%$ 이 된다.

비율(rate)은 특정 현상의 변동 속도를 표시하는 측도로, 단위시간 당 사건 수로 표시한다.

[예시] 위암 수술환자 100명 중 위암수술 후 만 1년이 되는 시점까지 재발한 사람이 40명이다.

- 위암 수술환자에서 재발환자 1명에 대해 재발하지 않은 환자는 몇 명인가(비)?

$$60 : 40 = 1.5:1$$

- 위암 수술환자에서 재발률(분율)은?

$$40 / 100 = 0.4$$

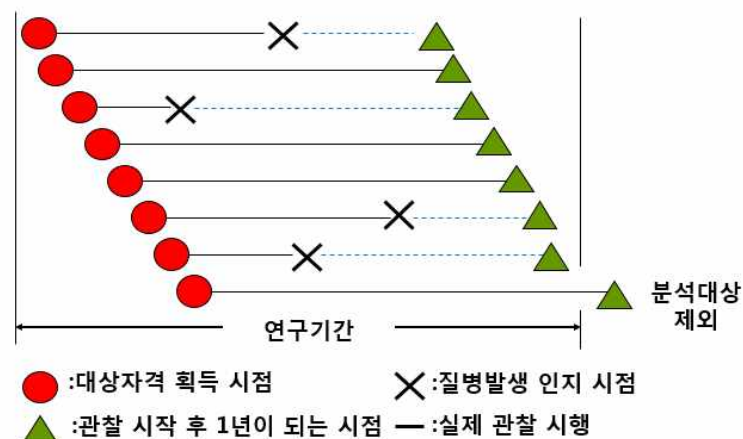
- 위암 수술환자에서 재발비율(비율)은?

$$40 / (60+20: \text{재발하지 않는 } 60\text{명은 } 1\text{년, 재발한 사람은 } 0.5\text{년으로 계산}) \\ = 0.5 / \text{year}$$

3.2. 발생률, 유병률, 사망률

발생률(incidence)과 유병률(prevalence)은 역학연구에서 질병의 빈도를 제시할 때 널리 사용되는 역학 지표이다. 발생률은 기본적으로 질병발생의 위험성이 있는 인구 집단에서 특정 기간 동안 새롭게 발생한 사건 수를 의미하며, 발생률을 산출하기 위해서는 전체 인구집단에서 해당 질병에 이환되어 있는 유병환자를 제외하고 추적관찰하는 것이 필요하다. 분모의 형태에 따라 분율(누적 발생률, cumulative incidence)과 비율(평균 발생률, incidence density)로 측정할 수 있다.

누적 발생률은 관찰 기간 동안 다른 이유로 중도에 탈락되지 않고 모두 특정기간 동안 관찰이 가능한 경우에 발생한 사건 수를 관찰 인구집단의 수로 나눠주어 산출하는 것으로, 1년 발생률, 3년 발생률과 같이 관찰 기간을 명시해야 한다. 누적발생률은 코호트 발생률 (cohort incidence rate), 발생 위험도(risk of incidence), 발생 분율 (incidence proportion) 등과 동의어이다.



[그림 3-1] 질병 발생률

$$\text{질병의 1년 발생률} = 4 / 7 = 57 \%$$

$$1,000 \text{ 명당 누적 발생률} = \frac{\text{해당 기간에 질병이 발생한 인구 수}}{\text{해당 기간의 질병이 발생할 가능성이 있는 인구 수}} \times 1,000$$

발병률(attack rate)과 2차 발병률(secondary attack rate)은 rate이라는 표현 때문에 비율로 오해할 수 있으나 누적 발생률, 즉, 분율(proportion)의 한 형태이다. 발병률은 특정 질병의 발생이 한정된 기간에만 일어나는 경우에 산출하는 지표로 급성전염병의 원인균에 폭로된 후 일정기간 내에서만 발생하는 질환의 발생률이 이에 해당한다. 2차 발병률은 전염병 역학에서 주로 사용하는 누적 발생률로 해당 질병에 면역이 없는 사람들이 발단 환자와 접촉하여 그 질병이 발생하는 율이다. 이를 통해 전염성의 강도나 예방방법의 효과를 확인한다.

[예시] 4,399 명이 사는 마을에 원인불명의 괴질이 발병하여, 77 가구에서 총 115명의 환자가 발생하였다. 괴질발병가구 77 가구에 살고 있는 사람은 모두 434명이다.

— 이 지역사회에서 괴질의 발병률은?

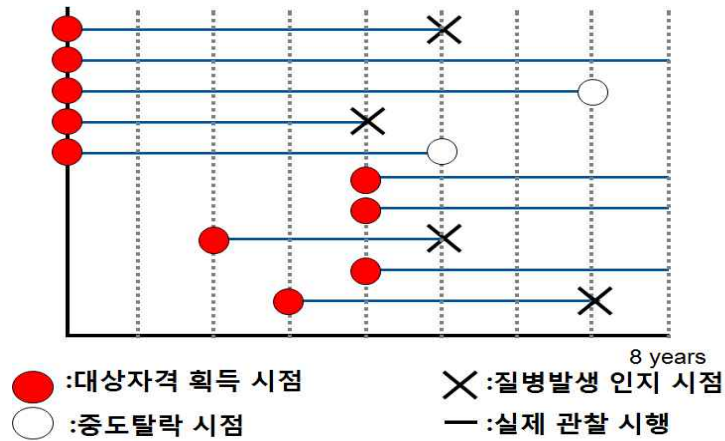
$$115 / 4,399 = 2.6\%$$

— 괴질 발생 가구당 한 명만이 1차 발병자(primary case, index case)라 가정한다면, 이들 가구에서의 2차 발병률은 ?

$$77\text{명} \rightarrow 115 \text{ 명} \rightarrow 434 \text{ 명}$$

$$\begin{aligned} &\text{추가 발병자 수} / \text{감수성 있는 가구원 수} \\ &(115-77) / (434-77) = 38 / 357 = 10.6\% \end{aligned}$$

평균 발생률은 발생한 사건수를 해당기간 동안 각 연구대상자의 관찰기간의 합(인년, person-year)으로 나눠주어 산출한 값으로, 분자는 누적 발생률과 같으나 분모가 다르다. 연구대상자마다 추적관찰기간이 일정하지 않은 들고나는 집단에서 발생률을 산출할 때에도 이용할 수 있다. 평균 발생률은 incidence density rate, person-time incidence 등과 동의어이다.



[그림 3-2] 질병 평균 발생률

$$\text{질병 발생률} = 4 / (5+8+7+4+5+4+4+3+4+4) * 1000 = 83.33/1,000 \text{ 인년 (person-years)}$$

$$1,000 \text{ 인년당 평균 발생률} = \frac{\text{해당 기간에 질병이 발생한 인구 수}}{\text{해당 기간의 질병이 발생할 가능성이 있는 각 대상자의 관찰기간의 합 (년)}} \times 1,000$$

유병률(prevalence)은 인구집단에서 질병 등 특정 상태를 가지고 있는 사람 수를 전체 인구집단의 수로 나눈 것으로 분율에 해당된다.

$$1,000 \text{ 명당 유병률} = \frac{\text{해당 기간에 질병 등 특정 상태를 가진 인구수}}{\text{해당 기간의 인구집단의 수}} \times 1,000$$

유병률은 시점 유병률 (point prevalence)와 기간 유병률 (period prevalence)로 제시할 수 있다. 시점 유병률은 인구집단에서 한 사람이 특정 시점에서 유병자일 확률을 의미한다.

특정 t 시점의 유병률은 t 시점의 유병자 수를 t 시점의 전체 대상자수로 나눠주어 산출한다. 기간 유병률은 인구집단에서 한 사람이 특정 기간 동안 어느 시점이던지 유병자일 확률을 의미하며, 분자는 관찰 시작 시점의 유병자수에 관찰 기간 동안 새로이 발생한 환자수를 더한 값이고, 분모는 관찰 기간 동안의 전체 대상자수이다.

[예시] 2009년 3월 1일부터 3월 31일까지 한 달간 특정 지역에 거주하는 전체 주

민 3,000명을 대상으로 흉부엑스선 촬영을 실시한 결과 60명의 결핵환자가 발견되었다. 그 가운데 3월 1일 이전에 결핵으로 진단받아 치료를 받고 있던 환자는 20명이었다.

- 누적 발생률 (3월 한달 간) - 40 명 / 2,980 명
* 3월 1일 이전 결핵 발생자 20명 제외
- 평균 발생률 (3월 한달 간) - 40 명 / 2,960 인-월 (person-month)
* 3월 1일 이전 결핵 발생자 20명 제외,
3월 1일 이전 결핵 발생자 20인-월 + 40명*0.5인-월 제외
- 시점 유병률 (2월 28일 기준) - 20 명 / 3,000 명
- 기간 유병률 (3월 한달 간) - 60 명 / 3,000 명

사망률 (mortality rate)은 행정통계의 기본이 되는 지표로 복지 및 보건의료 정책 수립을 위한 자료로 사용된다. 사망률은 주로 연 사망률 형태로 산출하며, 분자는 해당 년도에 사망한 인구수, 분모는 연앙인구 (해당 년도의 중앙인구)를 사용한다. 개 개인의 정확한 생존기간을 알기 어렵기 때문에 각 대상자의 추적관찰기간의 합의 근사치로서 연앙인구를 이용한다.

$$1,000 \text{ 명당 연 사망률} = \frac{\text{해당 년도에 사망한 인구수}}{\text{연앙인구 (해당 년도의 중앙인구)}} \times 1,000$$

치명률(case fatality rate)은 해당 질병에 이환된 사람 중에서 그 질병으로 인해 사망한 사람의 분율을 나타내는 지표로 해당 질병의 중증도를 가늠하는데 이용한다.

$$1,000\text{명당 치명률} = \frac{\text{특정 질병 발생 또는 진단 이후 특정 기간동안 사망한 인구수}}{\text{특정 질병이 있는 인구수}} \times 1,000$$

비례 사망률 (proportionate mortality rate)은 전체 사망자 중 특정 원인으로 사망한 사람수에 대한 분율이다. 우리나라에서 암의 비례사망률은 전체 사망자 중 암 사망

자수의 분율이다.

$$\text{암 비례사망률} = \frac{\text{2010년 암으로 인한 사망자 수}}{\text{2010년 사망자 수}}$$

3.3. 표준화법

발생률, 유병률, 사망률 등 역학 지표의 경우 서로 다른 인구집단에서 결과를 비교하거나, 한 인구집단에서 다른 기간에 산출된 결과를 비교하는데 이용된다.

<표 3-1> 볼티모어시의 이종 연령별 사망률

인종	전체 조사망률	연령별 조사망률 (인구 1,000명당)					
		<1	1-4	5-17	18-44	45-64	>=65
백인	14.3	23.9	0.7	0.4	2.5	15.2	69.3
흑인	10.2	31.3	1.6	0.6	4.8	22.6	75.9

1965년 볼티모어시의 인종별 사망률을 나타내는 위 표에 따르면 백인의 조사망률이 흑인에 비해 높은 것을 볼 수 있었는데, 이는 해당 시기의 생활여건, 의료환경이 백인에 비해 열악하였음을 고려할 때 해석하기 쉽지 않은 결과였다. 하지만, 각 연령별 조사망률을 보았을 때 일관적으로 흑인에서 높게 나타났으며 전체 조사망률에 따른 판단보다는 인구구조를 고려한 사망률에 대한 평가가 필요한 것을 확인할 수 있었다.

위의 예시처럼, 연령, 성별 등 역학 지표에 영향을 주는 특성이 각 인구집단에 따라 분포가 달라 조사망률 등 보정하지 않은 값을 이용하는데 한계가 있다. 연령 등 중요한 특성을 두 군에서 동일하도록 지표를 비교하는 표준화법 (standardized method)을 적용할 수 있다.

□ 직접표준화법 (direct standardized method)

두 집단의 사망률을 연령 표준화 한다고 할 경우, 직접표준화법의 경우 표준화하고자 하는 두 집단의 각 연령구간 별 사망률, 표준 인구집단의 연령구간 별 인구수에 대한 정보가 필요하다. 연령에 대한 직접표준화의 단계는 다음과 같다.

1) 연령별 인구수를 알고 있는 표준집단을 선정함, 2) 비교하고자 하는 두 집단 (A,

B 집단)의 연령별 율을 표준인구에 적용하여 표준 인구에서의 연령별 기대빈도수를 산출함, 3)각 집단의 연령별 기대빈도수를 합하여 표준집단의 총수로 나눔.

[예제] 직접 표준화 방법

<표 3-2> A, B 지역의 X 질병의 1,000명 당 조 발생률

	A지역	B지역
20-44	2.2	4.0
45-64	2.6	4.4
65-99	7.1	5.3
조발생률	4.2	4.6

심사평가원 자료를 이용하여 분석한 A, B 지역의 X 질병의 1,000명당 조 발생률이 각각 4.2, 4.6이었다. 이를 2010년 통계청 인구총조사 자료를 이용하여 연령 표준화하여 비교하고자 한다.

1) 연령별 인구수를 알고 있는 표준집단을 선정한다.

<표 3-3> 연령별 인구수를 알고 있는 표준집단을 선정

연령	인구1,000명당 발생률		인구총조사
	A지역	B지역	
20-44	2.2	4.0	18,520,287
45-64	2.6	4.4	12,820,420
65-99	7.1	5.3	5,422,832
조발생률	4.2	4.6	
총 인구수			36,763,539

2) 표준 인구에서의 연령별 기대빈도수를 산출한다.

<표 3-4> 표준 인구에서의 연령별 기대빈도수를 산출

연령	인구1,000명당 발생률		인구총조사	기대빈도수	
	A지역	B지역		A지역	B지역
20-44	2.2	4.0	18,520,287	2.2x0.001x 18,520,287=40,929	4.0x0.001x 18,520,287=74,716
45-64	2.6	4.4	12,820,420	2.6x0.001x 12,820,420=33,795	4.4x0.001x 12,820,420=56,403
65-99	7.1	5.3	5,422,832	7.1x0.001x 5,422,832=38,541	5.3x0.001x 5,422,832=29,010
조발생률	4.2	4.6			
총 인구수			36,763,539		

3) 연령별 기대빈도수를 합하여 표준집단의 총수로 나눈다.

<표 3-5> 연령별 기대빈도수를 합하여 표준집단의 총수로 나눔

연령	인구1,000명당 발생률		인구총조사	기대빈도수	
	A지역	B지역		A지역	B지역
조발생률 총 인구수 기대 빈도수의 합 연령 보정 발생률	4.2	4.6	36,763,539	113,265 113,265/36,763,539 x1,000=3.1	160,129 160,129/36,763,539 x1,000=4.4

□ 간접표준화법 (indirect standardized method)

간접표준화법은 연령군 별 지표를 모르고 사망 또는 발생자의 총 수만 아는 경우에 주로 사용된다. 또한, 특정 집단에서의 사망건수가 매우 적어서 연령별 산출이 불안정하여 직접표준화법의 적용이 어려운 경우에도 적용한다. 연령에 대한 간접표준화의 단계는 다음과 같다.

- 1) 표준인구집단의 연령별 율 (표준율)을 선정한다.
- 2) 비교하고자 하는 집단 (A 집단)의 연령별 총수에 연령별 율을 곱하여 기대빈도수를 산출한다.
- 3) A집단에서 실제 관찰된 빈도 총수를 기대빈도수의 합으로 나누어 표준화 (발생, 유병, 사망)비를 산출한다.
- 4) 표준인구집단의 전체 율 (crude rate)에 표준화 비를 곱하여 보정된 표준화 율 (age-adjusted rate)을 산출한다.

[예제] 간접 표준화 방법

1950년 미국에서 백인남자광부 534,533명 중, 결핵으로 436명이 사망한 것으로 집계되었다. 백인남자광부의 결핵 사망률이 일반인구집단의 백인남성보다 높은가?

전체 백인남성의 결핵 사망률은 10.9/1,000이었다.

- 1) 표준인구집단 (일반인구집단 중 백인남성)의 연령별 결핵 사망율 (표준율)을 선정한다.
- 2) 비교하고자 하는 집단 (백인남자광부)의 연령별 총수에 연령별 율을 곱하여 기대빈도수를 산출한다.

- 3) 백인 남자 광부 집단에서 실제 관찰된 빈도 총수를 기대빈도수의 합으로 나누어 표준화 사망비를 산출한다.

<표 3-6> 표준화 사망비 산출

연 령	1) 표준인구집단의 결핵 사망률 (100,000명 당)	2) 백인 남자 광부의 연령별 인구수	3) 기대빈도수
20-24	12.26	74,598	9.14
25-29	16.12	85,077	13.71
30-34	21.54	80,845	17.41
35-44	33.96	148,870	50.55
45-54	56.82	102,649	58.32
55-59	75.23	42,494	31.96
Total			181.09

표준화 사망비 (Standardized mortality ratio)

= Total observed deaths / Total expected deaths

= 436 / 181.09 * 100 = 241% (또는 2.41)

- 4) 표준인구집단의 전체 율 (crude rate)에 표준화 비를 곱하여 백인 남자 광부 집단의 보정된 표준화 율(age-adjusted rate)을 산출함.

Age-adjusted rate = 10.9/1,000 * 2.41 = 26.3/1,000

3.4. 상대위험도 (relative risk), 오즈비(odds ratio)

역학연구에서 요인-질병간의 관련성을 평가하기 위한 지표로 상대위험도(relative risk, RR)와 오즈비(odds ratio, OR)를 이용한다.

상대위험도 (relative risk)는 질병의 원인으로 가정한 특정 요인에 노출된 환자군과, 노출되지 않은 환자군으로 각각 인구집단을 구성하고, 각 집단에서 산출된 질병 발생률을 나눈 비(ratio)로 코호트 연구를 통해 산출되는 지표이다.

<표 3-7> 질병 발생 노출 유무

요인 노출	질병 발생(사례군)	비 발생(대조군)	추적관찰 기간의 합
노출됨	a	b	T1
노출되지 않음	c	d	T0

$$\text{상대위험도 (RR)} = \frac{\text{요인 노출군의 질병 발생율}}{\text{요인 비노출군의 질병 발생율}} = \frac{a / (a+b)}{c / (c+d)} = \frac{a / T_1}{c / T_0}$$

오즈비는 3.1장 대푯값, 분율, 비율에서 설명한 대응비(Odds)의 비로, 환자-대조군 연구에서 환자군의 요인 노출에 대한 대응비를 대조군의 요인 노출에 대한 대응비로 나눈 값이다.

$$\text{요인노출에 대한 오즈비 (OR)} = \frac{\text{환자군에서 요인 노출에 대한 오즈}}{\text{대조군에서 요인 노출에 대한 오즈}} = \frac{a / c}{b / d} = \frac{a \times d}{b \times c}$$

이는 요인노출에 따라 군을 나뉜 질병발생에 대한 대응비를 나타낸 값과 동일하며, 오즈비는 코호트 연구, 환자-대조군 연구에서 모두 사용할 수 있는 지표이다.

$$\text{질병발생에 대한 오즈비 (OR)} = \frac{\text{요인 노출군에서 질병 발생에 대한 오즈}}{\text{요인 비노출군에서 질병 발생에 대한 오즈}} = \frac{a / b}{c / d} = \frac{a \times d}{b \times c}$$

4. 연구설계

임상진료 현장에서의 진단과 치료에 있어서의 판단, 보건의료 정책 수립에 있어서 정책의 수립과 추진의 판단은 근거를 필요로 한다. 이러한 근거들은 다양한 연구들에서 구할 수 있는데, 의학연구는 사람을 직접 대상으로 하는지 여부에 따라 임상연구와 전임상연구로 나눌 수 있다. 임상연구는 사람을 대상으로 하는 연구로서 실험실적 연구나 동물을 대상으로 하는 연구들보다 높은 근거 수준을 가지며, 환자증례보고, 환자군연구, 환자-대조군연구, 코호트 연구, 임상시험으로 갈수록 점점 높은 근거의 위계를 갖는다.



[그림 4-1] 의학연구 설계와 근거의 위계

환자증례보고(Case Report)와 환자군연구(Case Series Study)는 가설을 생성하는 단계의 임상연구이다. 환자증례보고는 기존에 알려져 있지 않은 특이한 질병양상을 보인 환자의 특성과 치료에 따른 임상경과 및 의심되는 원인에 관하여 보고하는 연구로서 통계분석을 통한 정량적인 근거 생산을 위한 과정이 포함되지 않는다. 하지만, 환자증례보고는 질병과 치료에 대한 통찰을 제공하고, 이로부터 다양한 가설과 연구가 파생되는 의학연구의 첫 단계로서 진료와 연구의 가교역할을 담당하고 있다. 본 교재를 통하여 전달하고자 하는 건강보험 자료를 통한 근거생성에는 부합하지 않는 연구설계로 생각된다. 하지만, 독자들이 추후 설계하고 분석할 연구들의 시작이 환자증례보고에서 출발한 경우가 많을 것이며, 환자증례보고 문헌의 내용을 확인하면 건강보험 청구자료에서 확인할 수 없는 변수들에 대한 정보를 얻고 추후 자료연계를 통한 연구에서 어떤 변수들을 더 보완하여야 하는지에 대한 정보를 구할 수 있을 것이다.

환자군연구는 동일한 질병을 가진 환자들을 모아서 특성을 요약하고, 주요 특성에

다른 임상경과를 비교함으로써 질병의 원인에 관한 정량적 가설을 수립하는 후향적 연구설계이다. 환자증례보고가 사과를 맛보는 단계의 연구라면, 환자군연구는 사과를 분류하고 특성을 세어보는 연구설계라고 할 수 있다. 임상에서는 주로 어떤 질병을 진단받고 이미 치료를 받았던 환자들을 병원의무기록을 이용하여 후향적으로 찾아낸 후 환자들의 공통점을 파악하거나, 과거에 치료받았던 내용들을 비교하여 치료방법들 간 치료효과의 차이에 관한 가설을 제기하는데 활용하고 있다. 후향적이라는 특성을 고려할 때 단 시간에 효율적으로 질환에 대한 연구를 수행할 수는 있고, 특정한 환자들의 특성에 따라 어떤 질환, 또는 약물의 사용 양태를 정량적으로 기술하는 것 자체가 목적일 때에는 유용할 수 있으나, 대조군이 없기 때문에 환자증례보고를 통하여 질병발생의 원인에 대한 근거가 확립되기는 어렵다. 하지만, 분석적인 관찰연구를 계획할 때, 연구대상수 산출 등 필수적인 정량적인 가설이 환자군 연구를 통하여 만들어질 수 있기 때문에, 많은 경우 인과관계 정립의 과정에서 중간단계로 활용되고 있다.

위 그림의 근거위계 피라미드에서 가장 꼭대기에는 임상시험이 있지만, 건강보험 청구자료를 활용할 때 쓸 수 있는 설계가 아니므로 이를 논외로 할 때, 환자-대조군 연구와 코호트연구가 건강보험 청구자료를 활용하여 가장 상위의 진료와 정책의 근거를 만들어 낼 수 연구설계가 된다. 이 두 연구설계 중에서도 코호트연구가 환자-대조군연구에 비하여 상위에 있는데, 이는 비뚤림(bias)의 발생을 방지할 수 있는 방법론적인 수단이 코호트연구설계에 더 잘 확보되어 있고 코호트연구만의 장점이 있기 때문이다. 따라서, 환자-대조군연구를 통하여 노출요인과 질환사이의 관련성이 밝혀진 경우에도 코호트연구를 수행하여 부가적으로 더 많은 근거를 얻을 수 있다. 임산부의 DES (diethylstilbestrol) 노출에 따른 질선암 관련성은 암이 잘 발생하지 않는 소아 청소년기 여성에서 질선암(vaginal adenocarcinoma) 사례 보고에서 시작되었다. 비슷한 시기에 진단된 젊은 여성의 질선암 사례 수집을 통하여 DES노출이라는 공통점 파악하고 가설을 생성하였고, 환자군 8명과 대조군 32명을 모집한 환자-대조군연구를 통하여 DES와 질선암의 관련성 입증하였다.⁵⁾ 같은 사실을 코호트연구를 통하여 증명하고자 1940~1960년 대 임신 중 DES를 복용한 여성의 딸 3,000 여명을 대상으로 코호트를 구축하고 4년간 추적관찰하였으나 1979년까지 질선암 증례 발견하지 못하였다.⁶⁾ 이에 DES 노출군 4,536명과 DES 비노출군 1,544명으로 코호트를 확장하고 추적관찰기간을 1994년까지 늘려서 DES 노출군에서만 질선암 3례 발생 확인하여

5) Herbst AL, Ulfelder H, Poskanzer DC, Longo LD. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. 1971. American journal of obstetrics and gynecology. 1999;181(6):1574-5.

6) Hoover RN, Hyer M, Pfeiffer RM, Adam E, Bond B, Cheville AL, et al. Adverse health outcomes in women exposed in utero to diethylstilbestrol. The New England journal of medicine. 2011;365(14):1304-14.

유의한 관련성과 함께 노출에 따른 발생율을 제시하였다.⁷⁾ 본 장에서는 환자증례보고와 환자군연구를 통하여 생성된 가설을 확증하기 위한 분석적 관찰연구 설계의 특징을 설명하고 실제 수행사례를 살펴보고자 한다.

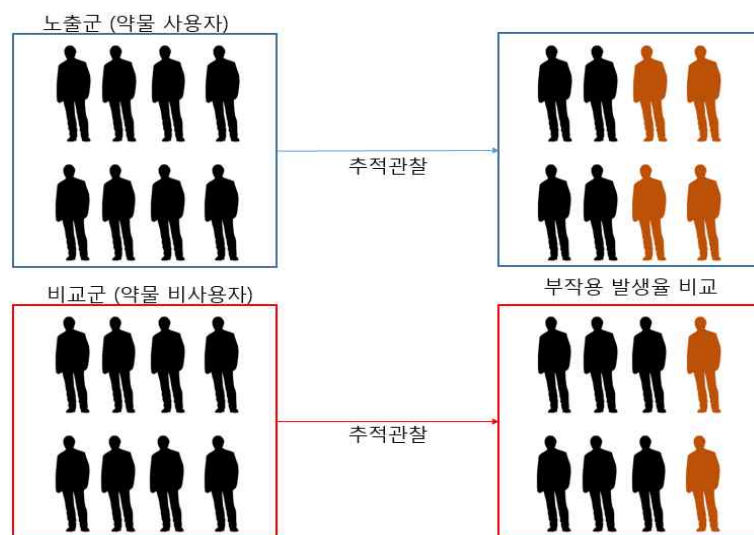
4.1. 코호트 연구

코호트연구는 연구대상이 되는 질환을 경험하지 않은 사람을 연구대상으로 선정한 다음, 질병발생의 원인으로 가정한 요인의 노출상태에 따라 연구대상 인구집단을 구성하고, 이들을 일정 기간 동안 추적관찰하여 특정 질병의 발생여부를 확인하여 요인의 노출과 질병 발생의 인과적 연관을 밝히는 연구방법이다 [그림 4-2]. 대상의 선정과 추적관찰기간, 그리고 발생율지표에 따라서 분류하면, 모든 구성원이 연구시작 시점에 한꺼번에 선정되고 일정기간을 추적관찰 후 같은 시점에 관찰을 종료하여 누적 발생률(Cumulative Incidence)을 발생율의 지표로 활용하는 고정 코호트(Fixed cohort)와 대상 선정시점, 종료시점이 다르고 따라서 추적관찰기간이 제각기 다르기 때문에 평균발생률(Incidence Density)을 발생율의 지표로 이용하는 들고나는 코호트(Dynamic cohort)로 나눌 수 있다. 전자는 사고로 인한 유해물질에 대한 대규모 폭로가 이루어진 상황에서 그 건강영향을 평가하는 연구로 적절하며, 후자는 대부분의 지역사회 기반의 질병 코호트연구들에 해당하는 형태라 하겠다. 또한, 질병이 발생하기 전 대상을 모집한 후 추적 관찰하여 질병발생을 확인하는 경우를 전향적 코호트연구, 질병이 발생한 시점에 대상자들을 모집하여 이전의 노출정보를 활용하여 노출여부에 따른 위험을 분석하는 코호트연구를 후향적 코호트연구라 한다.

건강보험자료를 이용하면 고정 코호트 형태의 분석도 가능하고, 들고나는 코호트 형태의 분석도 수행이 가능하다. 또한, 전향적 코호트연구도 가능하고 후향적 코호트연구도 가능하다. 하지만, 많은 경우 건강보험 청구자료의 장점을 가장 잘 살리는 접근이라 할 수 있는 후향적 코호트연구를 수행하는 경우가 많으며, 질병에 대한 코호트분석이기 때문에 들고나는 코호트 분석이 주로 수행된다. 장기간의 자료에서 관찰하고자 하는 코호트 구성원이 되려면 질병발생 가능성(Disease potential)이 있어야 한다. 이는 연구하고자 하는 질환이 걸려있지 않음을 의미하기도 하며, 해당질환이 걸릴 수 있는 가능성이 있어야 한다는 뜻이다. 가령 위절제를 받은 사람이 위출혈을 관찰하는 코호트연구의 대상이 될 수 없으며, 여성은 전립선암의 발생을 관찰하는 코호트연구의 대상자가 될 수 없을 것이다. 건강보험 청구자료를 활용하면 이러한 조건들이 쉽게 확인되어 선정제외기준으로 작용하여 코호트연구 대상자를 선정할 수

7) Hatch EE, Palmer JR, Titus-Ernstoff L, Noller KL, Kaufman RH, Mittendorf R, et al. Cancer risk in women exposed to diethylstilbestrol in utero. *Jama*. 1998;280(7):630-4.

있게 된다. 또한 요인 폭로정보(Exposure data) 수집이 진단 및 처방 데이터를 활용하여 효율적으로 이루어 질 수 있다. 따라서 입적된 대상자들을 요인 노출군과 비노출군으로 나누고, 추적관찰하여 질병 발생률을 산출하고, 군간의 발생율을 비교하여 위험요인 폭로와 질병발생 간 관련성 평가할 수 있게 된다.



[그림 4-2] 코호트연구의 구조의 흐름

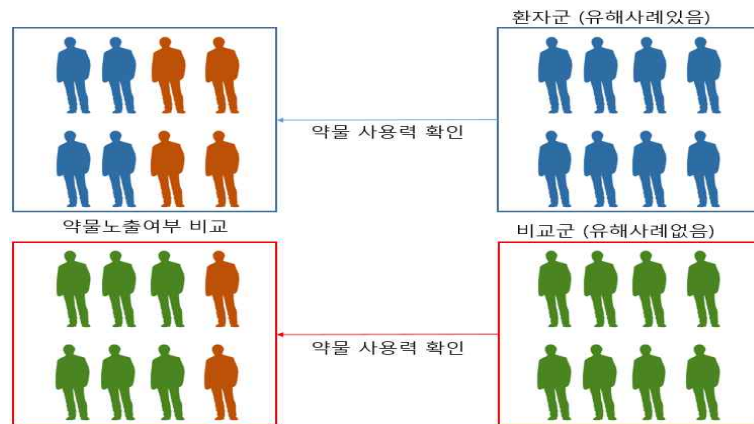
코호트연구의 장점은 위험요인의 노출과 관련하여 두드러진다. 위험요인에 관한 정보를 연구대상(질병)이 발생하기 전에 파악하기 때문에 환자-대조군연구처럼 환자의 기억에 의존하지 않으므로 더욱 정확하게 파악할 수 있으며, 동시에 위험요인과 질병 발생간의 시간적인 선후관계가 명확하기 때문에 인과관계의 평가에 더욱 유리하다. 특히, 질병발생 전의 생화학적 자료 수집이 가능하며, 요인 노출 자료를 후향적 혹은 간접적으로 획득하는 경우 신빙성이 의문시될 때 유용한 설계라 할 수 있다. 또한, 요인폭로에 대한 반복조사 혹은 측정이 가능하다. 노출요인에 대한 타당한 측정 자료를 활용하여 특정 요인의 노출로 인하여 발생할 수 있는 다양한 건강결과를 함께 파악할 수 있으며, 일반 인구집단에서처럼 특정 위험요인에 대한 노출빈도가 높지 않은 경우에도 충분한 검정력을 갖도록 설계하여 적용할 수 있다. 코호트연구를 수행하면 위험요인 노출에 따른 질병발생의 상대위험도뿐만 아니라 절대위험도, 기여위험도도 산출이 가능하여 임상적 판단뿐만 아니라 정책적인 판단을 돕는 근거로 활용할 수 있다. 하지만, 결과변수 확인과 관련하여 많은 단점을 갖고 있는 것도 사실이다. 코호트연구는 발생률이 낮은 질병과 특정 위험요인간의 관련성을 밝히고자 하는 경우에는 대규모의 연구대상을 필요로 하여 연구수행 가능성이 떨어진다. 특정 위험요인에 노출된 사람들을 대상으로 발생률이 낮은 장기간에 걸친 추적

관찰을 통하여 관찰하려면, 연구대상자뿐 아니라 연구자의 중도탈락이 문제가 되며, 추적관찰 기간 중도탈락이 많거나 선택적으로 특정 집단에서 일어나는 경우 비뚤림이 심각해진다. 또한 추적 관찰하는 기간 동안에 연구대상 위험요인에 대한 노출량과 강도가 변하는 것을 정확하게 측정하는 것은 여러 차례 방문과 측정이 필요하여 실제 수행하는 것이 매우 힘들며, 건강결과의 발생 여부를 파악하는데 많은 시간과 비용이 필요하다. 후향적 연구의 경우 시간이 단축되나 관련 자료의 과거 기록이 편중 혹은 일부만 있을 경우 비뚤림이 발생할 소지가 있다는 점들은 코호트연구의 단점이다.

건강보험 청구자료를 이용한 코호트연구는 이러한 단점들로부터 비교적 자유롭다고 할 수 있다. 하지만, 현재의 건강보험자료는 생활습관과 관련된 인자, 주요 임상검사의 결과 및 질병의 중등도와 관련된 정보가 부재하기 때문에 측정되지 않은 교란요인으로 인한 문제가 단점으로 지적되고 있다. 본 연구의 수행은 자료연계를 통한 분석의 활성화를 가져올 것이고, 결과적으로 이러한 문제를 극복할 수 있는 방안을 찾을 것으로 기대할 수 있다.

4.2. 환자-대조군연구

환자-대조군연구는 특정 질병을 가진 환자군과 그런 질병을 가지지 않은 대조군을 선정한 후, 과거의 특정 요인 노출 정도를 두 군 사이에 비교를 통하여 요인과 질병 사이의 인과적 연관성을 밝히는 연구설계이다 [그림 4-3]. 여기서 환자(case)라 함은 연구과정에 연구진이 확인할 수 있는 조건들에 입각하여 명백히 정의되는 것으로, 임상진료에서 확진을 위하여 활용하는 검사를 연구에서는 활용하지 않을 수도 있기 때문에 임상적인 의미의 환자와는 차이가 있을 수 있다. 이러한 차이의 정도는 연구에 할당되는 자원, 연구 대상자들이 환자인지 여부를 확인하는 세팅, 연구에 할당된 시간 등에 따라 달라질 수 있다. 계획단계에서는 연구결과의 활용을 고려하며 연구수행 여건상실무적으로 정의가 가능한 환자의 정의를 수립하는 것이 중요한 첫 단계가 된다. 건강보험 청구자료에 환자는 주로 진단명과 처방약물에 기반하여 이루어진다. 하지만, 경우에 따라서는 ‘질환으로 인하여 입원치료를 받은 환자’와 같이 특정한 임상적 상황을 만족하는 환자로 정의될 수 있으며, 보험제도에서 부여하는 산정 특례 기준을 만족하여 해당 코드를 갖고 있는 환자 등으로 정의될 수도 있다.



[그림 4-3] 환자-대조군연구 설계의 구조와 흐름

질병의 원인을 밝히는 환자-대조군연구를 설계할 때 환자는 새로 질병이 발생한 환자로 정의하는 것이 바람직하다. 만약, 새로 질병이 발병하지 않은 경우를 포함하여 환자를 정의하였다고 가정하여 보자. 분석 결과 특정 인자가 이 질병과 연관이 있다고 하였을 때, 이 인자가 질병을 일으키는 요인일 수도 있으나 질병이 걸리고 나서 환자가 생존하도록 해주는 요인일 수도 있다. 따라서 연구의 목적이 명확히 질병 발생의 원인을 확인하고자 할 때는 환자를 새로 질병이 생긴 환자를 대상으로 하는 것이 바람직하다. 이렇게 확인된 환자군에 비교할 대조군을 정의하고 양군에서 폭로의 정도를 비교하는 것이 환자-대조군연구라고 하였다. 코호트연구에서는 발생율을 구하고 군간의 발생율비로 연관성의 강도를 제시한다고 하였는데, 환자-대조군연구는 오즈비 (odds ratio)로 연관성의 강도를 제시한다. 노출의 확률비로 오즈를 구하고, 오즈의 비로 연관성을 나타내기 때문에 환자-대조군연구의 결과가 타당한 것이 되기 위해서는 환자군은 연구대상 질환자들을 대표할 수 있어야 하고, 대조군은 비질환자들을 대표할 수 있어야 한다. 또한, 환자군은 대조군과 필요한 정보의 수집 과정이 동일해야 한다. 가령, 환자군에 대해서 조사원이 더 꼼꼼히 요인 폭로정보를 수집하게 된다면, 또는 환자군의 대상자가 자신의 질병의 원인에 대한 관심으로 인하여 더 상세한 요인 폭로정보 제공을 하는 경향이 있게 된다면 연구 결과는 비뚤릴 수 밖에 없게 된다. 이러한 상황을 방지하게 위하여 조사원과 연구대상자들을 연구가설에 대해서 눈가림이 되도록 하며, 환자군과 대조군에서 같은 자료 수집도구를 활용할 수 있도록 준비한다. 연구가설의 군간 비교를 쉽게 하기 위해서는 연구계획 단계에서 교란변수 분포가 환자군과 대조군에 동일하게 만들어 주는 것이 바람직하다. 환자-대조군연구의 어려움으로 적절한 대조군을 선정하는 것을 들 수 있는데, 이러한 교란변수의 분포와 대조군 선정의 어려움을 동시에 해결해 주는 방법으로 짝짓기를 꼽을 수 있다. 이상적으로는 모든 교란변수가 균형을 이루는 것을 생각할 수도 있으나, 짝짓기의 조건을 너무 엄격하게 하면 짝을 찾지 못하고 탈락하게 되는 대상

자가 너무 많아지기 때문에 대부분 분석에서는 성별, 연령 5세 간격, 그리고 주된 질환 등 2-4개 정도의 변수에서 짝짓기를 하게 된다. 짝짓기에는 환자군 전체의 교란 변수의 분포가 대조군 전체의 교란변수의 분포와 동일하도록 하는 빈도 짝짓기 (frequency matching), 그리고 환자군 한 명에 대조군을 하나 혹은 둘 이상씩 짝짓는 개별 짝짓기(individual matching)가 있다. 짝짓기의 비율은 경험적으로 비용 노력에 비하여 검정력의 증가정도를 고려하여 1:4정도를 권하고 있지만 건강보험 청구자료를 이용할 때는 환자 1명에 대조군 4명을 초과하는 짝짓기도 종종 이루어지곤 한다.

코호트 연구에 비해 대상 수의 규모가 작아 작은 경비로 빠른 시간에 결론을 도출할 수 있다는 것은 환자-대조군연구 연구의 가장 큰 장점이다. 비교적 드물게 발생하는 건강결과와 관련된 위험요인을 파악하는데 유리하고, 단일 질병에 대해 여러 가지 위험요인을 동시에 검정할 수 있고, 비교적 희귀한 질환이나 긴 잠복기를 가지는 질병의 연구에 적용하여 효율적으로 결론을 도출할 수 있다. 따라서 앞서 살펴보았던 DES와 질선암의 예에서와 같이 보건학적으로 중요한 문제에 대해서 빠른 답을 필요로 할 때 가장 먼저 고려하여야 할 연구설계이다. 하지만, 환자-대조군연구는 위험요인에 노출된 사람이 매우 적은 경우에는 연구대상 또한 노출이 많지 않을 것이고 결과적으로 환자군과 대조군 모두에서 노출이 없을 것이기 때문에 적용하기 어렵다. 과거에 노출된 위험요인에 관한 정확한 정보를 파악하는 것은 쉽지 않아서 정보 비뚤림이 발생하기 쉽고, 원인의 폭로 정도가 매우 낮은 질환인 경우 폭로자가 환자군과 대조군 모두 0이 될 수도 있기 때문에 이런 경우 수행하다고 하여도 결론에 도달할 수 없다. 또한 특정한 위험요인을 대상으로 하여 위험요인에 노출된 사람에서 특정 건강결과의 발생을 관찰하는 연구가 아니고, 이미 건강결과가 발생한 환자로부터 과거에 경험한 위험요인에 관한 정보를 파악하는 연구이기 때문에 특정 요인에 따른 건강결과의 발생률을 직접 산출할 수는 없다. 하지만, 무엇보다 어려운 점은 적절한 대조군을 선정하는 것과 관련된 문제이다. 건강보험 청구자료를 이용할 때는 질환이 없는 사람 전체 목록을 이용한 무작위추출 또는 짝짓기가 가능하기 때문에 비교적 대조군 선정의 문제를 수월하게 진행할 수 있다. 하지만, 선택비뚤림의 가능성과 관련된 이슈는 코호트연구에서와 달리 환자-대조군연구에서 특히 문제가 된다. 이 때문에, 건강보험 청구자료를 활용한 후향적 코호트연구가 가능한 주제라면 환자-대조군연구보다는 코호트연구가 비뚤림의 가능성이라는 측면에 있어서 우선적으로 고려될 수 있다. 하지만, 건강보험 청구자료를 이용하여 코호트 보다는 환자-대조군 연구설계가 더 적절한 상황들이 있다. 가령, 신생아기 감염과 약물사용에 따른 소아기 질병 발생을 파악하는 코호트의 구축은 신생아기 연령자료의 부정확성으로 인하여 쉽지 않기 때문에, 결과적으로 질환 발생 소아와 대조군 소아를 비교하는 환자-대조군연구 설계가 가능할 것이다.

5. 청구자료를 이용한 성과연구⁸⁾

성과연구(outcomes research)는 실제 인구집단에서 이루어지는 일상진료(routine care) 환경에서 노출(exposure) 또는 치료(treatment)의 성과를 평가하는 연구이다. 이때 성과는 질환이나 치료로 인한 사망과 같은 최종 결과뿐만 아니라 임상적 결과(증상, 실험실적 검사치 등), 신체적 또는 정신적 기능과 삶의 질(quality of life), 치료의 결과에 대한 만족도 등의 환자보고성과(patient-reported outcomes)와 의료비용까지 포함한다. 이런 성과연구를 통해 임상진료지침을 개발하거나, 진료의 질을 평가하고 치료의 효과를 파악할 수 있다⁹⁾.

성과연구에서 관심있는 약물 또는 의료기술의 임상적 효과를 비교하기 위해 근거의 생성이 필요한 경우 근거 생성을 위한 자료원이 필요하다. 자료원은 크게 일차자료원과 이차자료원으로 구분할 수 있는데, 일차자료원이란 연구자가 직접 무작위대정임상시험을 수행하거나, 환자 등록자료를 구축하는 등 연구목적에 맞게 수집한 자료원을 의미한다. 일차 자료원의 구축이 불가능할 경우 다른 목적으로 기 구축된 이차자료원, 예를 들어 병원 의무기록자료, 청구자료 등을 이용하여 임상적 효과를 비교할 수 있다.

청구자료를 이용한 성과연구는 후향적 연구로 수행되는데, 일반적으로 후향적 데이터베이스를 이용한 연구에서는 RCT에 비해 일상진료(routine care)를 반영할 수 있고, 대규모 환자를 장기간 추적관찰하여 임상효과를 파악할 수 있으며, 비교적 짧은 시간에 최소한의 비용으로 연구를 수행할 수 있어서 시의적절한 연구결과의 도출이 필요한 경우에도 유용하다.

청구자료를 이용하여 성과연구를 수행하고자 할 경우, 관심있는 약물 또는 의료기술이 모두 급여대상이며, 임상적 효과를 비교할 수 있는 의미있는 결과변수가 청구자료에서 정의 가능해야 한다. 이 때 비교하고자 하는 약물 또는 의료기술간의 비교가능성을 확보하는 것이 무엇보다 중요하다. 따라서 임상적 효과를 비교할 때 바이어스를 최소화하기 위해 인과성(causal effect)을 판단할 수 있는 코호트 연구 또는 환자-대조군 연구 등의 잘 설계된 연구설계 방법을 적용해야 하며, 청구자료에서 파악할 수 있는 관찰된 교란요인들(confounders)을 보정하기 위해 적절한 통계분석 방법을 적용해야 한다.

청구자료를 이용한 성과연구 수행 이전에 연구설계 및 자료분석방법 등을 포함한 연구 프로토콜을 작성하여 수행할 것을 권장한다. 연구 프로토콜 작성 시 고려해야

8) 배은영, 김윤희, 이태진, 이혜재, 장은진. 의료기술 경제성 및 급여적정성 평가 기준 개선 연구. 건강보험심사평가원. 2015.

9) 장은진, 안정훈, 정선영, 황진섭, 이자연, 심정임. 측정된 교란요인을 고려한 성과분석 방법. 한국보건의료연구원. 2013.

하는 사항은 다음과 같다.

5.1. 연구설계

약물 또는 의료기술의 임상적 효과를 평가하기 위해 연구설계 방법을 결정할 때는 연구가설과 청구자료의 특성뿐만 아니라 연구설계 방법의 강점과 제한점을 함께 고려하여 해당 연구설계 방법을 결정한 근거를 설명해야 한다. 자세한 사항은 4장. 연구설계를 참고한다.

□ 코호트 연구

청구자료를 이용한 코호트 연구는 후향적 코호트 연구에 해당되며, 분석대상 치료법과 비교대상 치료법의 적응증을 가지는 연구대상자를 코호트로 먼저 정의해야 한다. 이때 코호트를 정의할 입적 기간(index period)과 과거력을 정의할 과거 기간(history period), 결과발생을 파악할 추적관찰기간(follow-up period)을 사전에 정해야 하는데, 이는 과거력 변수의 특성과 결과 변수의 임상적인 특성을 고려하여 정한다. 예를 들어 사용가능한 자료가 5년 자료인 경우, 결과 변수가 발생할 때까지의 추적관찰기간을 길게 정의해야 하는 경우 추적관찰기간을 과거 기간이나 입적 기간에 비해 길게 정의해야 하며, 장기간 파악이 필요한 과거력이 있는 경우 입적 기간이나 추적관찰기간에 비해 과거 기간을 길게 정의해야 한다. 이는 선행연구 또는 전문가 의견 등을 반영하여 결정하고 이에 대한 근거를 설명해야 한다.



[그림 5-1] 코호트 연구

코호트 대상자의 코호트 입적일(index date)은 입적 기간 내에 해당 치료법이 시행된 날짜로 정의해야 하지만, 청구자료에서는 동일 명세서 내의 세부 청구내역에 대한 날짜 정보를 파악할 수 없기 때문에 일반적으로 치료법에 대한 청구가 처음으로 발생한 명세서의 요양개시일자로 정의한다. 하지만 전문가 의견 등에 따라 치료법

시행 일자에 대해 조작적으로 정의(예를 들어 입원 하루 후 등)할 수 있는 경우에는 이에 대한 근거를 설명하고 정의에 따라 입적 날짜를 정의한다.

추적관찰기간은 코호트 입적 날짜 이후부터 추적종료시점까지로 정의하며, 추적 종료시점은 치료법과 결과변수의 특성에 따라 결정한다.

□ 환자-대조군 연구

청구자료를 이용한 환자-대조군 연구는 코호트내 환자-대조군 연구(nested case-control study)로 설계가 가능한데, 이는 먼저 연구대상자를 코호트로 정의하고, 정의된 코호트 내에서 결과가 발생한 사람을 환자군으로 정의하고, 결과가 발생하지 않고 환자군과 유사한 대조군을 사전에 정의된 코호트 내에서 선정하는 방법이다.



[그림 5-2] 코호트 내 환자-대조군 연구

코호트 내 환자-대조군 연구에서도 코호트 연구와 동일하게 입적 기간, 과거 기간, 추적관찰기간 및 입적일을 정의하고, 추적관찰기간내 처음으로 결과가 발생하는 경우 환자군으로 정의하고 이 때 명세서의 요양개시일을 결과발생일(event date)로 정의한다.

환자군과 유사한 대조군을 선정하기 위하여 개별 매칭(individual matching) 방법을 사용할 수 있으며, 환자군과 대조군의 매칭비율은 보통 1:1로 선정하는 것이 일반적이나, 통계적 검정력을 증가시키기 위해 1:4까지 매칭을 실시하기도 한다.

□ 연구설계 결정 시 고려할 점

관심 결과 변수가 드물게 일어나는 경우 코호트 연구 설계를 이용할 경우에는 발생건수가 작아서 인과성을 파악하는데 비뚤림이 생길 가능성이 높다. 따라서 이런 경우 환자-대조군 연구 설계를 고려할 수 있다.

코호트 연구에서 치료법과 결과변수 간의 연관성(association)을 나타내는 측도로 상대위험도를 사용할 수 있는 반면에, 환자-대조군 연구에서는 상대위험도를 계산할 수 없다. 따라서 환자-대조군 연구에서는 상대위험도 대신 오즈비를 이용하여 연관성을 나타낸다. 하지만 결과 발생이 드문 경우 오즈비는 상대위험도와 근사적으로 같아지므로, 관심 결과 변수가 드물게 일어나서 환자-대조군 연구설계를 이용하였을 경우 오즈비를 상대위험도와 같이 해석할 수 있다.

5.2. 연구대상자 선정

청구자료를 대상으로 분석대상 치료법과 비교대상 치료법의 적응증을 가지는 연구대상자를 선정하기 위해서 선정/제외 기준(inclusion/exclusion criteria)을 정의한다. 먼저 입적 기간 내에 치료법에 대한 청구가 있는 대상자를 선정하고, 사전에 정한 제외기준에 해당하는 대상자를 제외하여 치료법의 적응증을 가지는 연구대상자로 제한(restriction) 한다. 이때 치료법의 관심 결과를 과거에 경험한 대상자를 제외하여 치료법에 대한 신환자를 연구대상자로 정의하고자 할 경우 과거력 기간에 해당 치료법에 대한 청구가 있었던 경우를 과거력으로 정의하여 제외한다. 연구대상자의 선정 및 제외 기준은 적응증의 특성에 따라 상병코드 또는 처치코드, 약물코드 등으로 정의하고, 필요한 경우 이들 코드들의 조합으로 정의할 수도 있다.

5.3. 변수 정의

분석대상 치료법 및 비교대상 치료법, 결과변수, 교란요인을 정의하기 위하여 상병코드, 약물코드, 처치코드, 재료코드 등을 이용하여 조작적 정의가 필요하다. 이를 위해 선행연구 및 전문가 의견을 참고할 수 있으며, 근거를 설명해야 한다. 결과변수가 병원내 사망인 경우 진료결과변수 또는 상병코드를 이용하여 정의할 수 있으나, 의료기관이 아닌 장소에서 사망한 경우는 청구자료로 파악할 수 없기 때문에 전체 사망 또는 특정 상병으로 인한 사망을 정의하기 어렵다. 추적관찰 종료 시점은 연구가설 및 결과변수 특성에 따라 마지막 방문일 또는 가용한 청구자료의 마지막 일자가 될 수 있다.

청구자료를 이용하여 변수를 정의할 경우 분류 비뚤림(classification bias)이 발생할 수 있는데, 포괄수가제 또는 급여제한 등으로 실제로 해당 변수에 대한 경험이 있는데도 경험이 없는 것으로 오분류할 수 있으며, 약물의 경우 처방받았지만 복용하지 않았을 경우 해당 변수에 대한 경험이 없는데도 경험이 있는 것으로 오분류할 가능성이 있다.

5.4. 통계분석¹⁰⁾

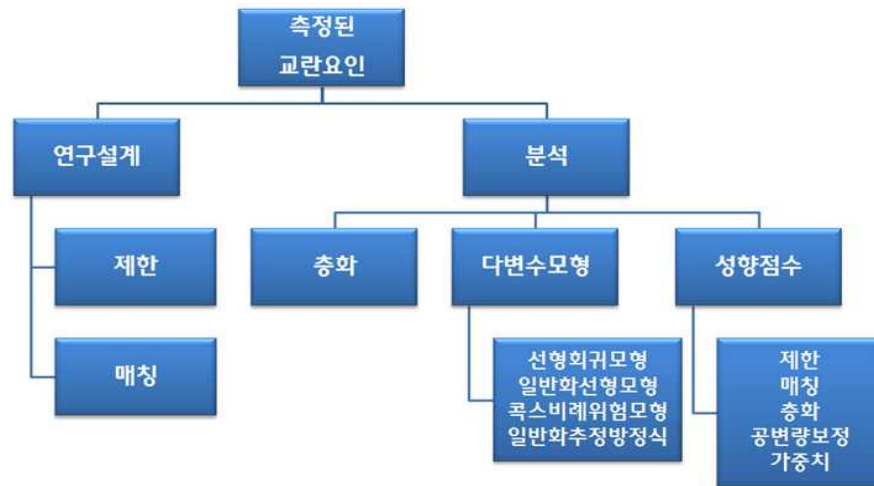
청구자료를 이용하여 관심있는 의약품, 치료법 등의 임상적 효과를 비교하기 위해 연구가설 및 연구설계에 적합한 방법론을 적용해야 한다. 후향적 코호트 연구 또는 환자-대조군 연구와 같이 관찰연구 형태로 수행되는 연구에서는 비뚤림이 발생할 가능성이 높는데, 이들 비뚤림 중 치료와 결과변수 간의 관련성을 왜곡시키는 비뚤림을 교란(confounding)이라고 한다. 그리고 교란요인은 치료와 관련성이 있으면서 결과에 영향을 주는 제3의 요인으로 정의할 수 있다. 따라서 치료법의 선택과 관련성이 있으면서 결과에 영향을 주는 교란요인을 청구자료를 이용하여 정의하고, 정의된 측정된 교란요인을 통제하는 분석 방법을 적용해야 한다.

측정된 교란요인을 연구 설계 측면에서 통제하는 방법은 제한과 매칭이 있다. 제한 방법은 연구에 포함되는 대상자들의 동질성을 확보하기 위하여 적절한 선정/제외 기준을 이용하여 연구대상자들을 제한하는 방법이며, 매칭 방법은 비교하고자 하는 두 군이 비슷한 특성을 가지도록 기준이 되는 군과 교란요인의 분포가 동일하거나 유사하도록 비교군을 선정하는 방법이다.

측정된 교란요인을 분석 측면에서 통제하는 방법은 층화, 다변수 모형(multivariable model), 성향점수(propensity score) 방법이 있다. 층화분석은 결과에 영향을 줄 수 있는 교란요인이나 효과변경인자를 다양한 범주별로 층을 나누어 층 내에서 동일한 분석을 시행하는 방법이다. 교란요인의 수준별로 층을 나누어 각각 분석하는 방법으로 적용이 간편하며 교란요인의 수가 많지 않을 때 효과적인 방법이다.

다변수 모형은 교란요인을 설명변수로 다변수모형에 포함시켜 이를 보정하는 방법이며, 성향점수 방법은 연구 대상 군들의 관찰된 교란요인들의 균형을 맞추어 두 그룹간의 유사성을 확보하기 위해 사용할 수 있는 방법이다. 성향점수 방법은 다차원(multi-dimensional) 공변량들을 성향점수라는 일차원(one-dimensional) 점수로 차원을 축소하는 방법이며, 추정된 성향점수는 제한, 매칭, 층화, 공변량 보정, 가중치(weight) 방법을 적용할 수 있다. 연구가설과 자료가 주어졌을 때, 측정된 교란요인을 통제하기 위한 방법을 적용할 수 있는 상황을 요약하면, 제한, 매칭, 층화 방법은 교란요인의 수가 적을 경우 적절한 방법이며, 만일 고려해야 하는 교란요인의 수가 많을 경우에는 교란요인들을 성향점수로 요약하여 매칭 또는 층화 방법을 적용하는 것이 더 적절하다. 이들 방법론에 대한 자세한 적용방법은 장은진 등(2013)을 참고하기 바란다.

10) 장은진, 안정훈, 정선영, 황진섭, 이자연, 심정임. 측정된 교란요인을 고려한 성과분석 방법. 한국보건의료연구원. 2013.



[그림 5-3] 측정된 교란요인을 고려한 분석방법

연구가설과 자료가 주어졌을 때, 측정된 교란요인을 통제하기 위한 방법에 따라 적용할 수 있는 상황을 요약하면 다음 표와 같다. 제한, 매칭, 층화 방법은 교란요인의 수가 적을 경우 적절한 방법이며, 만일 고려해야 하는 교란요인의 수가 많을 경우에는 교란요인들을 성향점수로 요약하여 매칭 또는 층화 방법을 적용하는 것이 더 적절하다.

<표 5-1> 측정된 교란요인의 통제방법 선정

적용 가능한 상황	
설계적 측면	
제한	특성이 동질한 연구대상자를 선정하여 치료효과 추정
매칭	교란요인의 분포가 동일하거나 비슷한 대상자를 매칭하여 치료효과 추정
분석적 측면	
층화	교란요인 또는 효과변경인자 수준별로 치료효과 추정
다변수 분석	교란요인을 보정하여 치료효과 추정
성향점수	비교군들간 균형을 맞추어 치료효과 추정

5.5. 결과 해석 시 고려할 점

청구자료를 이용한 연구 결과 해석시 연구설계 측면에서의 잠재적인 제한점을 고려하여 비뚤림의 발생 가능성에 대해 고찰할 필요가 있다. 또한 청구자료를 이용할 경우 다른 임상자료에 비해 표본수가 굉장히 클 수 있으므로 이런 경우 대표본 효과 (large sample size effect)로 통계적으로 유의한 결과가 나올 가능성이 높다. 따라서

이런 경우 통계적 유의성 뿐만 아니라 임상적 유의성도 함께 고려하여 결과 해석이 필요하다. 아울러 청구자료에서의 연구 결과의 일반화 가능성에 대해서도 고찰할 필요가 있다.

5.6. 이차자료원을 이용한 연구 체크리스트

STROBE(Strengthening the Reporting of Observational studies in Epidemiology)는 세 가지 주요 연구 설계-코호트연구, 환자-대조군연구, 횡단면 연구에 대한 관찰연구를 보고하기 위한 가이드라인으로 개발되었다. STROBE 가이드라인은 제목, 요약, 소개, 방법, 결과 및 토론 섹션과 관련된 22개의 항목으로 구성된다. STROBE 가이드라인은 의학 저널 편집자의 국제위원회에서 생물 의학 저널에 제출된 원고에 대한 통일 요건을 언급하고 있으며, BMJ Open, British Medical Journal, Lancet은 원고 제출 지침서에 STROBE 가이드라인을 참고한다.

최근 청구자료와 같이 일상진료상황을 반영한 건강데이터를 이용한 연구수행이 증가면서, STROBE와 같은 기존 지침에서 다루지 않은 문제를 해결하기 위하여 STROBE 지침을 확장하여 RECORD(REporting of studies Conducted using Observational Routinely collected health Data) 지침이 개발되었다¹¹⁾. RECORD지침은 STROBE의 확장판이므로, STROBE항목과 비교하여 정리할 수 있는데, 구체적인 체크리스트는 부록을 참고한다. Benchimol EI 등(2015)에는 체크리스트 뿐만 아니라, 각 부분별로 잘 보고된 연구사례에 대한 예시도 제시하고 있으므로, 청구자료를 이용한 연구 수행시 참고하기 바란다.

이차자료원을 이용한 연구를 수행하고자 할 경우 연구윤리 측면에서 연구자는 연구윤리위원회의 승인을 받아야 하며, 연구진의 전문성을 확보해야 한다. 다시 말해 사용된 자료원에 대한 지식과 이차자료 연구 수행에 대한 경험이 있는 연구자 및 통계학, 의료 및 건강연구에 대한 배경 지식을 가지고 있으며 방법론 및 통계 전문지식을 제공할 수 있는 생물통계학자가 필요하다.

추가적으로 연구에 활용한 데이터에 대한 품질을 설명하기 위해서 다음과 같은 지표를 사용할 수 있다¹²⁾.

11) Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. (2015) The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med 12(10): e1001885. doi:10.1371/journal.pmed.1001885

12) Iron, K & Manuel, D. (2007). Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data. Online. ICES.

1. 데이터가 올바른가?
 - 정확성 : 데이터가 진실을 반영하는가?
2. 데이터가 완전한가?
 - 완전성 : 데이터에 수집된 모든 기록이 포함되는가?
 - 포괄성과 적용 범위 : 데이터가 목표한 연구대상자의 100%를 포함하는가?
3. 데이터를 신뢰할 수 있는가?
 - 신뢰성 : 데이터가 재현 가능한가?
 - 유효성 : 데이터가 의미가 있는가?
4. 데이터를 사용할 수 있는가?
 - 익명성 : 데이터가 사생활 보호, 절차 및 관행을 준수하는가?
 - 연결가능성 : 건강 관리 시스템의 복잡성을 반영하기 위하여 다른 데이터와 연결 할 수 있는가?
 - 적시 : 데이터 수집과 사용 사이에 약간의 시간차가 있는가?
 - 유용성 : 데이터를 쉽게 구성하고 액세스 할 수 있으며 쉽게 사용할 수 있는 형식으로 제공되는가?

6. 사례

6.1. 개요

본 사례에서는 류마티스 관절염 (Rheumatoid Arthritis; RA)환자에서 전통적인 RA 치료제 (Disease Modifying Anti-Rheumatic Drug; DMARD)와 생물학적인 DMARD인 종양괴사인자-알파 억제제 (Tumor necrosis factor- α inhibitor; TNFi)에 사용에 따른 고관절 골절의 발생율을 비교하는 후향적 코호트 자료 구축과 분석의 과정을 소개한다. 본 사례는 kim et al (2012)¹³⁾을 참고로 하여 청구자료를 이용한 후향적 코호트연구의 수행에 관한 이해를 돕기 위한 사례로 작성되었다. 본 사례에서 사용된 연구설계와 변수 정의가 참고논문과 차이가 있는 부분들이 있으나 우리 자료의 특성을 반영하기 위한 것이었으며, 이에 임상적 해석을 위한 자료의 보완없이 본 자료만으로 치료제와 부작용간의 인과적 연관성으로 참고논문과 같은 의미를 부여하는 것은 가능하지 않음을 미리 밝힌다.

6.2. 자료원

건강보험심사평가원의 건강보험 청구자료에서 진료일 기준으로 2007년 1월 1일부터 2015년 12월 31일까지 건강보험 및 의료급여로 상급종합병원, 종합병원, 병원, 요양병원, 의원, 정신요양병원, 보건소, 보건지소, 그리고 보건의료원을 방문한 RA 환자의 전체 의료이용자료를 이용하였다. RA 환자는 상병코드로 혈청양성 류마티스 관절염 (Seropositive rheumatoid arthritis, ICD-10 code: M05), 그 밖의 류마티스 관절염 (Other rheumatoid arthritis, ICD-10 code: M06)을 가지는 경우로 정의하였다. 이때 서면청구 또는 포괄수가제로 청구되거나 추가로 청구된 명세서는 제외하였으며, 연구에 활용한 건강보험 청구자료 내역은 다음 표와 같다.

13) Effects of disease-modifying antirheumatic drugs on nonvertebral fracture risk in rheumatoid arthritis: a population-based cohort study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2012;27(4):789-96. Epub 2011/12/14.

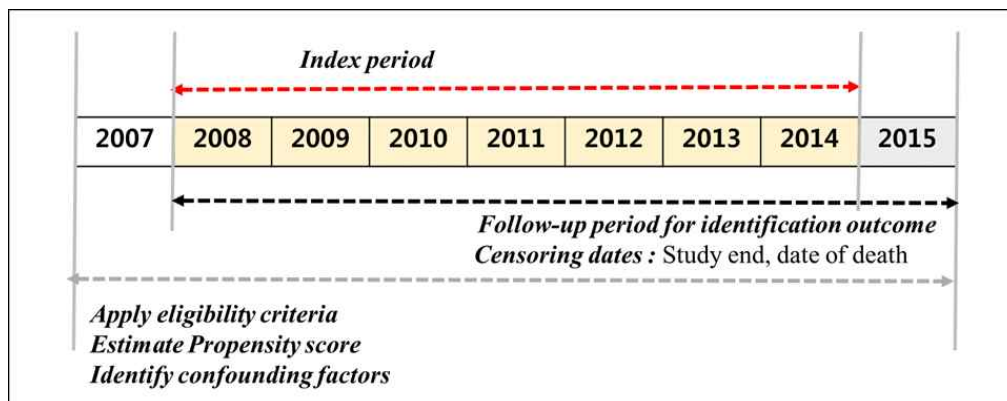
<표 6-1> 건강보험 청구자료 내역

구분	내용	
진료기간	2007.01.01 ~ 2015.12.31 (9개년도)	
요양기관 종별	상급종합병원, 종합병원, 병원, 요양병원, 의원, 정신요양병원, 보건소, 보건지소, 보건의료원	
보험자 종별	건강보험, 의료급여	
대상자	RA 환자	
청구자료 상세내역	테이블	내용
	T200 명세서일반내역	명세서 요약정보
	T300 진료내역	세부처치 및 원내 처방내역
	T400 수진자 상병내역	상병정보
	T530 원외처방전 상세내역	외래 처방 약제 정보

6.3. 연구설계

□ 연구대상자 선정

RA 환자에서 생물학적인 류마티스 치료제 (Biologic DMARD; bDMARD)와 전통적인 치료제 (traditional DMARD; tDMARD) 사용에 따른 고관절 골절의 발생을 비교하기 위해 건강보험심사평가원의 건강보험 청구자료를 활용하여 후향적 코호트 연구 (retrospective cohort study)를 시행하였다.



[그림 6-1] Schematic description of the study period

2007년 1월 1일 부터 2015년 12월 31일 사이에 RA 주상병 수진자로서 DMARD 처방을 받은 환자 중, 2008년 1월 1일부터 2014년 12월 31일 사이에 DMARD를 처음 처방받은 환자를 연구대상자로 선정하였다. 이 선정 작업을 위하여 RA 진단하에 추출된 전체 자료에서 DMARD 처방이 처음으로 나온 명세서의 요양개시일자를 코호트 입적일로 정의하였다. 선정기준과 제외기준은 다음 표와 같다. 2014년 12월 31일까지

로 입적기간을 한정한 것은 적어도 1년 이상의 관찰기간을 확보하여 투약에 따른 골절 결과를 확인하기 위한 장치이다.

<표 6-2> 연구대상자 선정/제외 기준

구분	내용
선정기준	<ul style="list-style-type: none"> • 2007-2015 RA 환자 중 2008-2014 기간 중 처음으로 DMARD 처방이 확인된 환자
제외기준	<ul style="list-style-type: none"> • tDMARD로 monotherapy 처방만 이루어진 환자 • 전체 코호트 기간 중 종양, 간부전, 신부전, 후천성 면역결핍증의 진단으로 인한 청구가 있는 환자 • 장기 이식을 시행 받은 환자 • cyclophosphamide를 투여받은 환자 • 코호트 입적일 이전에 대퇴골절로 인한 청구가 있는 환자

코호트 입적일을 기준으로 이전 1년까지 자료를 이용해 대상자의 질환 과거력 등을 파악하고, 추적관찰기간을 코호트 입적일 이후 2015년 12월 31일 이내 추적종료시점까지로 정의하였다. 추적종료시점은 고관절 골절로 인한 입원이 발생한 경우는 결과변수가 처음 발생한 시점으로 정의하고, 관심 결과가 발생하지 않은 경우는 2015년 12월 31일까지로 정의하였다.

□ 치료법 정의

DMARD 처방은 T300 진료내역, T530 원외처방전 상세내역에서 DMARD 코드의 유무로 정의하였다. 연구에서 고려한 DMARD의 목록은 다음과 같다.

<표 6-3> 연구에서 치료법의 정의를 위하여 고려한 DMARD의 목록

No	DMARD	ATC code
1	adalimumab	L04AB04
2	etanercept	L04AB01
3	infliximab	L04AB02
4	golimumab	L04AB06
5	mycophenolate mofetil	L04AA06
6	azathioprine	L04AX01
7	bucillamine	M01CC02
8	d-penicillamine	M01CC01
9	hydroxychloroquine	P01BA02
10	leflunomide	L04AA13
11	minocycline	A01AB23
12	mizoribine	L04AX
13	auranofin	M01CB03
14	sulfasalazine	A07EC01
15	thioguanine	L01BB03
16	methotrexate	L04AX03
17	cyclosporine	L04AD01
18	tacrolimus	D11AH02

후향적 코호트 연구 분석에서 DMARD 노출에 따른 결과변수 발생의 인과적 연관성을 보여주기 위한 용량반응관계, 기간반응관계를 제시하나 본 예시에서는 이러한 과정까지 제시하지는 않았다. 하지만, 위 약물 목록에서 제시하는 ATC 코드의 출처인 세계보건기구 홈페이지를 참고하여 각 약물의 Defined Daily Dose (DDD)를 파악할 수 있는데, 성인에서 해당 적응증에 대하여 1일 사용량으로 권장되는 용량인 DDD를 단위로 하여 누적사용량을 파악하고, 이 누적사용량에 따른 골절위험을 제시하는 방식으로 용량반응관계를 제시하고, 청구자료에 명시된 처방일자와 처방의 길이를 이용하여 골절 발생일시와의 인접여부를 판단하는 분석을 수행할 수 있다는 설명으로 이 부분은 같음하고자 한다.

□ 결과 발생 정의

bDMARD 군과 tDMARD 군의 고관절 골절 발생을 비교하기 위해, 연구대상자의 입적일 이후에 ICD 10 code로 S72.0 fracture of neck of the femur, 또는 S72.1 pertrochanteric fracture, 또는 S72.2 subtrochanteric fracture을 청구의 주진단으로 가지는 입원이 있는 경우 가지는 경우를 결과변수의 발생으로 정의하였다.

사망은 진료결과구분코드가 사망이거나, 상병코드가 급성 심장사로 기술된 것 (I461), 원인 미상의 기타 급사(R96), 순간적 사망(R960), 증상의 발생으로부터 24시간 이내에 일어난 달리 설명되지 않는 사망(R961), 지켜본 사람이 없었던 사망(R98), 기

타 불명확하고 상세불명의 사망원인(R99)인 경우로 정의하였다.

□ 공변량 정의

결과변수인 대퇴골절의 발생에 영향을 미치는 공변량으로 연령, 성별을 비롯하여 코호트 입적일 이전 1년 이내 발생한 골다공증 관련인자, 약물사용 관련인자, 그리고 의료이용 관련인자를 고려하였다. 세부적인 공변량 정의 기준은 다음표와 같다.

<표 6-4> 분석에서 고려한 공변량의 목록

구분	내용
Demographic factors	Age sex Osteoporosis diagnosis (M80, M81) Use of bisphosphonate
Osteoporosis related factors	Prior fracture at forearms, upper arms, pelvis or vertebra (S32, S42, S43) Parkinson's disease (G20) Demetia (F00-F03) Cumulative dose of oral glucocorticoids, mean (SD) Anticonvulsants Benzodiazepines
Medication use	SSRIs Proton pump inhibitors Beta blockers Opioids Number of all prescription drugs, mean (SD)
Health care utilization	Hospitalizations Number of physician visits, mean (SD)

여기서 약제 관련 요인들에 대해서는 세계보건기구 의약품 분류기준을 활용하여 각 약물군에 속하는 성분명을 확인하였다. 성분명코드는 건강보험심사평가원 성분명 마스터파일을 통하여 확인하여, 국내에서 보험 적용이 되는 약에 한하여 분석에 포함할 수 있었다.

<표 6-5> 공변량으로 고려된 약물의 목록

구분	내용
Bisphosphonate	etidronic acid , clodronic acid, pamidronic acid, alendronic acid , ibandronic acid, risedronic acid, zoledronic acid
Anticonvulsants	Barbiturates and derivatives methylphenobarbital, phenobarbital, primidone, barbexaclone, metharbital Hydantoin derivatives ethotoin, phenytoin, amino(diphenylhydantoin) valeric

		acid, mephenytoin, fosphenytoin paramethadione, trimethadione, ethadione
	Oxazolidine derivatives	
	Succinimide derivatives	ethosuximide, phensuximide, mesuximide
	Benzodiazepine derivatives	clonazepam
	Carboxamide derivatives	carbamazepine, oxcarbazepine, rufinamide, eslicarbazepine
	Fatty acid derivatives	valproic acid, valpromide, aminobutyric acid, vigabatrin, progabide, tiagabine
	Other antiepileptics	sultiame, phenacemide, lamotrigine, felbamate, topiramate, gabapentin, pheneturide, levetiracetam, zonisamide, pregabalin, stiripentol, lacosamide, carisbamate, retigabine, perampanel, brivaracetam, beclamide
	Natural opium alkaloids	morphine, opium, hydromorphone, nicomorphine, oxycodone, dihydrocodeine, papaveretum
	Phenylpiperidine derivatives	ketobemidone, pethidine, fentanyl
Opioids	Diphenylpropylamine derivatives	dextromoramide, piritramide, dextropropoxyphene, bezitramide, methadone
	Benzomorphan derivatives	pentazocine, phenazocine
	Oripavine derivatives	buprenorphine
	Morphinan derivatives	butorphanol, nalbuphine
	Other opioids	tilidine, tramadol, dezocine, meptazinol, tapentadol
Benzodiazepines	flurazepam, nitrazepam, flunitrazepam, estazolam, triazolam, lormetazepam, temazepam, midazolam, brotizolam, quazepam, lorazepam, doxefazepam, cinolazepam	
Selective serotonine reuptake inhibitors	zimeldine, fluoxetine, citalopram, paroxetine, sertraline, alaproclate, fluvoxamine, etoperidone, escitalopram	
Proton pump inhibitors	esomeprazole, pantoprazole, rabeprazole, omeprazole, lansoprazole	
Beta blockers	amosulalol, atenolol, betaxolol, bisoprolol, evantolol, carteolol, celiprolol, metoprolol, nebivolol	

스테로이드의 누적 노출량의 경우 각 스테로이드별로 효능이 다르고, 이에 따른 부작용의 정도도 다름이 알려져 있다. 이에, 모든 스테로이드를 하이드로코티손 등가 용량 (hydrocortisone equivalent dose)으로 변환하고 누적 노출량을 구하였으며, 누적 노출량 산출을 위한 등가용량 전환표는 다음과 같다.

<표 6-6> 공변량으로 고려된 부신피질호르몬과 하이드로코티존 등가용량

code	generic name	dose (mg)	Equivalent dose of hydrocortisone (mg)
316100BIJ	betamethasone	9	270
116502BIJ	betamethasone	4	120
296900ATB	betamethasone	0.25	7.5
116401ATB	betamethasone	0.5	15
116501ATB	betamethasone	0.5	15
141901ATB	dexamethasone	0.5	12.5
141903ATB	dexamethasone	0.75	18.75
141904ATB	dexamethasone	4	100
142001BIJ	dexamethasone	4	100
142201BIJ	dexamethasone	5	125
142202BIJ	dexamethasone	4.37	109.25
142203BIJ	dexamethasone	4	100
243201ATB	triamcinolone	1	5
243202ATB	triamcinolone	2	10
243203ATB	triamcinolone	4	20
243301BIJ	triamcinolone	200	1000
243303BIJ	triamcinolone	40	200
243305BIJ	triamcinolone	50	250
217001ATB	prednisolone	5	20
217003ASY	prednisolone	1	4
217004ASY	prednisolone	3	12
217101BIJ	prednisolone	110	440
217104BIJ	prednisolone	100	400
217301BIJ	prednisolone	1000	4000
217302BIJ	prednisolone	250	1000
217401ATB	prednisolone	6.65	26.6
193301ATB	methylprednisolone	16	80
193302ATB	methylprednisolone	4	20
193303ATB	methylprednisolone	8	40
193304ATB	methylprednisolone	2	10
193501BIJ	methylprednisolone	200	1000
193502BIJ	methylprednisolone	40	200
193601BIJ	methylprednisolone	125	625
193602BIJ	methylprednisolone	250	1250
193603BIJ	methylprednisolone	40	200
193604BIJ	methylprednisolone	500	2500
170901ATB	hydrocortisone	10	10
170905ATB	hydrocortisone	20	20
170906ATB	hydrocortisone	5	5
171201BIJ	hydrocortisone	100	100
171202BIJ	hydrocortisone	250	250
171203BIJ	hydrocortisone	500	500

□ 통계분석방법

연구대상자의 특성 중 범주형 변수는 빈도와 백분율로 제시하고, 연속형변수는 평

균과 표준편차로 나타내었다. bDMARD군과 tDMARD군의 기저 특성을 비교하기 위해 연속형 변수인 경우 독립 t-검정(independent t-test), 범주형 변수인 경우 카이제곱 검정(Chi-square test)을 사용하여 p-value를 산출하였다.

기저 불균형을 보정하기 위한 방법으로서 성향점수(propensity score)를 이용한 방법을 적용하였다. 이를 위해 대상자의 1년 이전 위험인자들을 이용하여 로지스틱 회귀모형을 이용하여 성향점수를 추정하고, 성향점수에 따라 Greedy algorithm을 이용하여 양군을 1:1로 개별매칭(individual matching)하였다.

매칭 전후의 공변량 균형을 검토하기 위해 표준화 차이(standardized difference, STD)를 사용하였고, 표준화 차이가 0.1보다 큰 경우 해당 공변량이 불균형이라고 판단하였다. 이분형 변수인 경우 표준화 차이는

$$d = \frac{(\hat{p}_{bDMARD} - \hat{p}_{tDMARD})}{\sqrt{\frac{\hat{p}_{bDMARD}(1 - \hat{p}_{bDMARD}) + \hat{p}_{tDMARD}(1 - \hat{p}_{tDMARD})}{2}}}$$

그리고 연속형 변수인 경우 표준화 차이는

$$d = \frac{(\bar{x}_{bDMARD} - \bar{x}_{tDMARD})}{\sqrt{\frac{(s_{bDMARD})^2 + (s_{tDMARD})^2}{2}}}$$

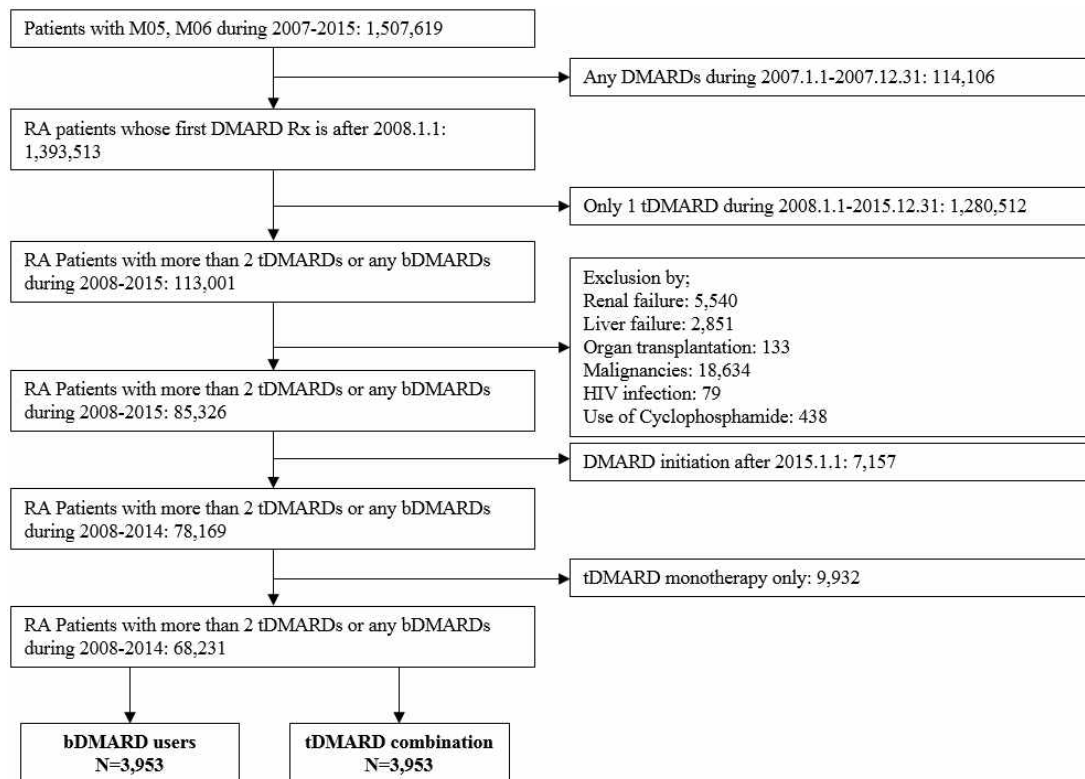
로 표현된다.

bDMARD군과 tDMARD군에 따른 고관절골절의 발생 위험은 콕스비례위험모형(cox proportional hazard model)으로 분석하였으며, 비례가정은 log-minus-log 생존그림을 이용하여 평가하였다. 위험비(hazard ratio, HR)는 95% 신뢰구간(confidence interval, CI)과 유의확률(p-value)을 함께 제시하였다. 고관절 골절의 경우 사망에 의하여 관찰이 불가능해 지는 결과변수이므로 경쟁위험(competing risk)을 고려한 통계분석 방법의 적용이 필요하나, 본 사례에서는 고려하지 않았다. 모든 통계분석은 SAS® Enterprise Guide version 6.1 (SAS Institute, Cary, NC, USA)을 사용하여 수행하였고, 통계적 유의성은 유의수준 5%로 판단하였으며 모든 검정은 양측으로 진행되었다.

6.4. 연구결과

□ 연구대상자 선정

연구 대상자 선정의 과정은 다음 그림과 같다. 최초 M05, M06 주진단 코드로 추출된 1,507,619 명의 RA환자에서 상기 선정제외기준을 적용한 결과 3,953명의 TNF inhibitor 사용자와 같은 수의 tDMARD 사용자를 짝지어 비교하게 되었다.



[그림 6-2] 연구대상자 선정의 과정

□ 연구대상자 기저특성

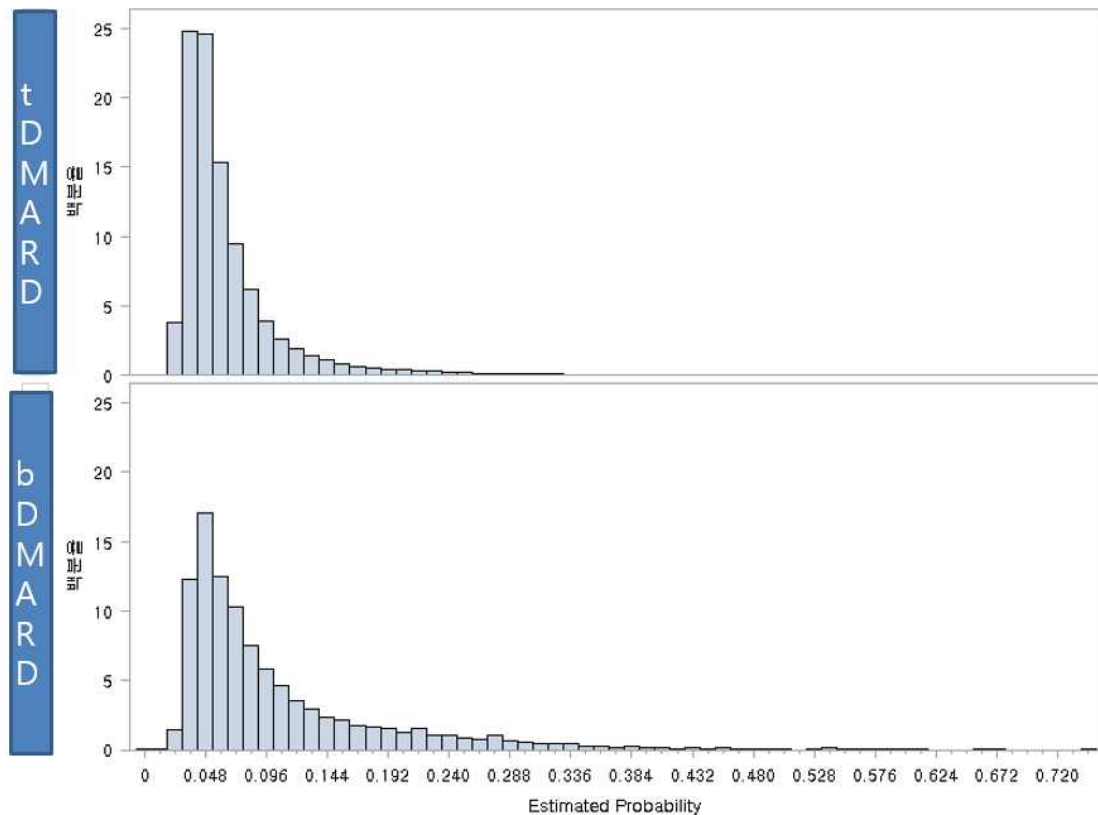
RA환자 중 tDMARD군과 bDMARD군의 기저특성을 살펴본 결과, proton pump inhibitor, opioids의 사용, 그리고 전체 방문수와 처방약의 개수를 제외한 대부분의 변수에서 통계적으로 유의한 차이가 있었다.

<표 6-7> 성향점수 짝짓기 전의 tDMARD군과 bDMARD군의 기저특성 비교

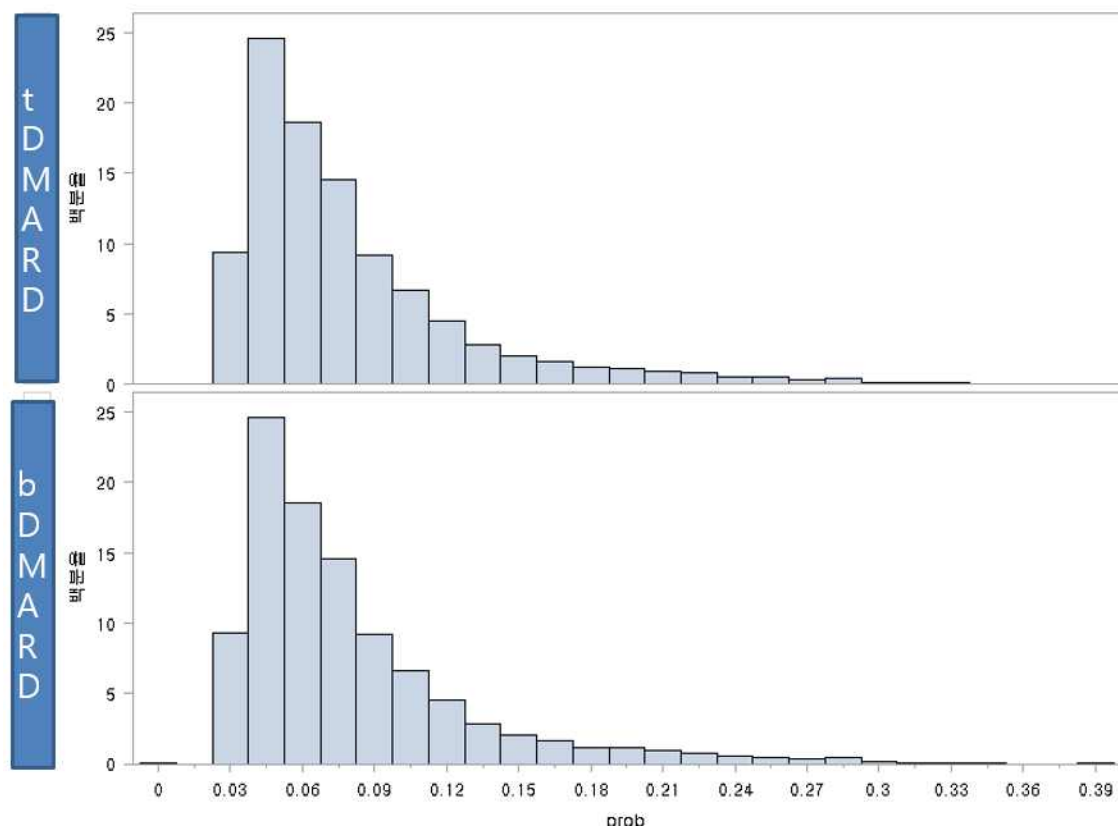
Characteristics		tDMARDs		bDMARDs		p-value
		n	%	n	%	
Demographic factors	Age, mean (SD)	51.3	13.7	43.4	16.2	<0.0001
	Male sex	13,390	21.1	1,696	36.2	<0.0001
	Osteoporosis diagnosis	16,759	26.4	1,016	21.7	<0.0001
Osteoporosis related factors	Use of Bisphosphonate	6,541	10.3	325	6.9	<0.0001
	Prior osteoporotic fractures	1,317	2.1	75	1.6	0.0278
	Parkinson's disease	167	0.3	5	0.1	0.0399
	Dementia	612	1.0	20	0.4	0.0002
Medication use	Cumulative dose of glucocorticoids, mean (SD)	1401.9	2562.9	1807.9	2770.1	<0.0001
	Anticonvulsants	6,406	10.1	398	8.5	0.0005

Health care utilization	Benzodiazepines	8,428	13.3	678	14.5	0.0183
	SSRIs	2,221	3.5	135	2.9	0.0268
	Proton pump inhibitors	17,263	27.2	1,296	27.7	0.4498
	Beta blockers	2,851	4.5	161	3.5	0.0007
	Opioids	50,962	80.2	3,743	79.9	0.6582
	Number of all prescription drugs, mean (SD)	39.4	21.9	39.4	22.0	0.9119
	Hospitalizations	19,433	30.6	1,207	25.8	<0.0001
	Number of physician visits, mean (SD)	31.4	29.0	31.9	28.7	0.24385

성향점수를 로지스틱 회귀모형을 이용하여 산출하여 그 분포를 보면, 아래 그림과 같이 두 군에서 불균형이 존재함을 알 수 있다.



bDMARD 환자 1명에 같은 성향점수를 갖는 환자 1명을 짝을 짓는 방법으로 양군의 균형을 맞추었고, 그 결과 양군의 성향점수의 분포는 아래의 그림과 같이 일치하게 되었다.



성향점수 짝짓기를 수행한 후, 성별을 제외한 모든 변수의 standardized difference는 0.1보다 작아서 두 군간 기저특성의 균형을 이루었다.

<표 6-8> 성향점수 짝짓기 이후 tDMARD군과 bDMARD군의 기저특성 비교

Characteristics		tDMARDs		bDMARDs		p-value
		n	%	n	%	
Demographic factors	Age, mean (SD)	46.6	14.5	47.0	14.2	-1.9×10^{-9}
	Male sex	843	21.3	1,342	34.0	-0.29
Osteoporosis related factors	Osteoporosis diagnosis	822	20.8	953	24.1	-0.07
	Use of Bisphosphonate	294	7.4	313	7.9	-0.04
	Prior osteoporotic fractures	69	1.8	67	1.7	0.00
	Parkinson's disease	10	0.3	5	0.1	0.04
	Dementia	29	0.7	20	0.5	0.03
	Cumulative dose of glucocorticoids, mean (SD)	1507.8	2392.3	1631.7	2475.7	-0.05
Medication use	Anticonvulsants	353	8.9	365	9.2	0.00
	Benzodiazepines	511	12.9	549	13.9	-0.03
	SSRIs	115	2.9	114	2.9	0.00
	Proton pump inhibitors	1,067	27.0	1,131	28.6	-0.04
	Beta blockers	136	3.4	157	4.0	-0.05
	Opioids	3,124	79.0	3,150	80.0	-0.02
Health care	Number of all prescription	38.2	21.9	39.5	22.5	-1.1×10^{-15}

utilization	drugs, mean (SD)					
	Hospitalizations	1,001	25.3	1,051	26.6	-0.05
	Number of physician visits, mean (SD)	29.5	28.2	32.8	22.4	-1.7*10 ⁻²⁰

□ 임상적 효과 분석

RA환자에서 tDMARD군과 bDMARD군에서 대퇴골절의 발생을 콕스비례위험모형으로 분석하였다. 기저특성에서 균형을 이루지 못한 성별은 보정변수로 추가하였다. 이때 tDMARD에 비한 bDMARD의 상대위험비는 1.40 배로 (0.76-2.57)로 위험증가의 경향은 있었으나 통계적으로 유의하지는 않았다 (p-value= 0.29).

<표 6-9> tDMARD와 bDMARD의 대퇴골절 발생의 비교

Treatment groups	HR	95% CI	p-value
tDMARD	Reference		
bDMARD by adjustment [†]	1.58	(1.16-2.16)	0.004
bDMARD by PSM*	1.40	(0.76-2.57)	0.29

[†]Adjusted for age, sex, osteoporosis, bisphosphonate, prior osteoporotic fracture, Parkinson's disease, dementia, cumulative dose of glucocorticoid, anticonvulsants, benzodiazepines, SSRIs, PPIs, beta blockers, opioids, number of all prescribed drugs, hospitalizations, number of physician visits

*Propensity score matching following adjustment for sex

□ 결론

본 사례에서는 RA 환자에서 tDMARD와 bDMARD의 투여에 따른 고관절 골절의 위험을 비교한 결과, tDMARD 군에 비해 bDMARD 군에서 위험이 높은 경향은 있으나 그 정도가 통계적으로 유의하지 않아서 결론을 내릴 수 없었다. 참고한 Kim et al.(2012)에서는 더 큰 환자집단을 대상으로 분석을 수행하였으나 통계적으로 유의한 결과를 관찰하지 못하였기 때문에 비슷한 경향의 결과로 보일 수도 있다. 하지만, 개요에서 설명한 한계로 인하여 엄밀한 임상적 해석과 참고논문의 결과와의 직접비교 보다는, 본 사례를 참고하여 실제의 분석을 구체적으로 계획하고 수행하는데 참고가 될 수 있기를 희망한다.

6.5. 프로그램 예시

□ 류마티스관절염 환자 선정

먼저 상병코드를 이용하여 류마티스관절염 환자를 정의하기 위하여 T40 테이블에서 M05, M06 상병코드를 가지는 행을 선택하고, 명세서 기준의 상병변수를 생성하기

위하여 SQL문을 사용한 후, 이를 명세서조인키 기준으로 T20 테이블에 조인한다. 그리고 최종적으로 류마티스관절염 상병을 가지는 명세서만을 선택한다.

```

/* T40 테이블에서 류마티스관절염 상병코드를 가지는 행 선택 */
data t40_RA;
  set hira.t40;
  if substr(SICK_CD,1,3) in ('M05', 'M06') then RA=1; /* 류마티스 관절염 */
run;

/* 각각의 명세서에 대한 류마티스관절염 상병변수 생성 */
proc sql;
  create table t40_RA1 as
  select spec_id_sno, sick_cd,
         max(RA) as ra
  from t40_ra
  group by spec_id_sno;
quit;

/* T20 테이블에 T40에서 생성한 류마티스관절염 상병변수 조인 */
proc sql;
  create table t2040_ra as
  select a.*, b.*
  from hira.t20 as a
  left join t40_RA1 as b
  on (a.spec_id_sno=b.spec_id_sno);
quit;

/* 류마티스 관절염(RA) 환자들에 대한 명세서만 선택 */
data t2040_ra1;
  set set t2040_ra
  where ra=1;
run;

```

위의 프로그램 결과 생성되는 테이블은 다음과 같다.

	SPEC_ID_SNO	ra	DMD_TP_CD	DMD_FRM_CD	FOM_TP_CD	JID	SEX_TP_CD	PAT_AGE
1	10000935	1 0	3		031	564944	1	41
2	10000979	1 0	3		031	1412621	1	47
3	10000980	1 0	3		031	1412621	1	48
4	10000981	1 0	3		031	1412621	1	48
5	10001417	1 0	3		031	103835	2	48
6	10001418	1 0	3		031	103835	2	48
7	10001529	1 0	3		031	145055	2	84
8	10003570	1 0	3		031	183036	2	89
9	10004473	1 0	3		031	359652	2	83
10	10005098	1 0	3		031	915555	2	40
11	10005099	1 0	3		031	915555	2	40
12	10005890	1 0	3		031	988990	1	73
13	10005891	1 0	3		031	1307912	2	62

※ 프로그램 Tip

① if substr(SICK_CD,1,3) in ('M05', 'M06') then RA=1

⇒ T40의 상병코드가 입력되어있는 'SICK_CD' 변수에서 substr함수를 통해 해당 질병에 해당하는 코드가 있을 경우 해당 질병 변수에 1이라는 값을 넣어 변수를 생성

SICK_CD	M0609	substr(SICK_CD, 1, 3) ⇒	M06	then RA = 1 ⇒	RA
	K602		K60		1
	M0590		M05		.
	M0642		M06		1
					1

② select spec_id_sno, max(RA) as ra
group by spec_id_sno

⇒ 동일한 명세서 조인키를 가지는 경우 RA 변수가 가지는 값들 중 최대값을 RA 변수에 입력하도록 하며, GROUP BY문을 이용해 동일한 여러 행(명세서)들을 모아 한줄로 요약

SPEC_ID_SNO	RA	group by SPEC_ID_SNO ⇒	SPEC_ID_SNO	RA
41958	1		9041	.
41958	1		41958	1
9041	.			
9041	.			

□ 치료법 정의

DMARD 처방은 T300 진료내역, T530 원외처방전 상세내역에서 DMARD 코드의 유무로 정의하였다. 이때 DMARD의 주성분명에 해당하는 주성분코드를 추출하기 위하여 약품코드, 주성분코드, 주성분명이 들어 있는 약가파일을 이용하여, SQL문의 like 함수를 사용하여 사전에 정의된 주성분명에 해당하는 행을 선택한다.

/*치료법의 정의를 위하여 고려한 DMARD 목록에서 일부만 선택*/

```
proc sql;
create table code.drug_master_RA as
select *
from drug_master
where
generic_nm like 'adalimumab%' or
generic_nm like 'etanercept%' or
generic_nm like 'infliximab%' or
generic_nm like 'golimumab%' ;
quit;
```

위에서 추출된 drug_master_ra에 포함되어 있는 주성분명 코드를 이용하여 T30 테이블과 T53 테이블에서 DMARD 약물을 사용한 명세서를 SQL문의 subquery 기능을 이용하여 다음과 같이 추출할 수 있다. 여기서는 총사용량을 이용하여 약물사용량을 정의하였다.

```

/*치료법 정의 */
/*T30테이블의 약품성분명 코드 중 일부 DMARD약물을 충족하는 명세서 추출*/
proc sql;
  create table replace_30 as
  select SPEC_ID_SNO,
  sum(TOT_USE_QTY_OR_EXEC_FQ) as ra_drug_t30 /* 동일 명세서 내 총사용량 합계 변수 생성*/
  from Hira.t30
  where GNL_NM_CD in (select generic_code from code.drug_master_ra )
                      /*DMARD약물을 충족하는 명세서 추출*/

  group by spec_id_sno;
quit;

/*T53테이블의 약품성분명 코드 중 일부 DMARD약물을 충족하는 명세서 추출*/
proc sql;
  create table replace_53 as
  select SPEC_ID_SNO,
  sum(TOT_USE_QTY_OR_EXEC_FQ) as ra_drug_t53 /* 동일 명세서 내 총사용량 합계 변수 생성*/
  from Hira.t53
  where GNL_NM_CD in (select generic_code from code.drug_master_ra )
                      /* DMARD약물을 충족하는 명세서 추출*/

  group by spec_id_sno;
quit;

/*위의 T30 T53테이블에서 추출한 명세서를 통합하여 DMARD약물을 투약한 명세서 추출*/
data drug_ra;
  merge replace_30 replace_53;
  by spec_id_sno;
/*T30 T53 테이블의 약물 투약량을 sum하여 DMARD약물 투약량 변수 생성*/
drug_ra_use = sum(ra_drug_t30, ra_drug_t53);
if drug_ra_use > 0 then DMARD=1;
                      /*DMARD약물 투약 실적이 있는 명세서 추출하여 RA치료법에 대한 테이블생성*/
  keep SPEC_ID_SNO DMARD;
run;

/*류마티스 (RA)환자에 대한 테이블에 DMARD변수 조인*/
proc sql;
  create table t2040_ra_dmard as
  select a.* , b.*
  from t2040_ra1 as a left join drug_ra as b
                      /*t2040_ra1: 류마티스(RA)환자 테이블, drug_ra: DMARD 변수 테이블*/
  on (a.spec_id_sno=b.spec_id_sno)
  group by jid;
quit;

```

위의 프로그램 결과 생성되는 테이블은 다음과 같다.

	SPEC_ID_SNO	ra	DMARD	RECU_FR_DD	DMD_TP_CD	DMD_FRM_CD	FOM_TP_CD	JID
1	12719509	1	.	20150803	0	3	021	1000011
2	5959303	1	.	20150411	0	3	031	1000013
3	24937272	1	.	20151113	0	3	031	1000049
4	22914403	1	.	20150609	0	3	031	1000049
5	22914402	1	.	20150603	0	3	031	1000049
6	23829321	1	.	20150930	0	3	031	1000049
7	20806042	1	.	20150713	0	3	031	1000049
8	20806043	1	.	20150724	0	3	031	1000049
9	12535212	1	.	20150305	0	3	031	1000049
10	12535213	1	.	20150318	0	3	031	1000049

※ 프로그램 Tip

① where

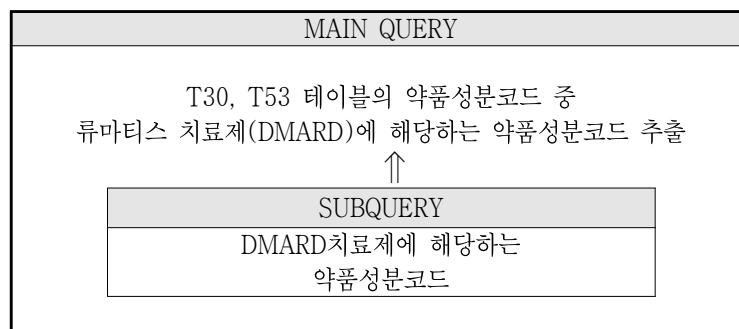
generic_nm like 'adalimumab%' or
generic_nm like 'etanercept%' or
generic_nm like 'infliximab%' or
generic_nm like 'golimumab%'

⇒ like 'A %' 는 특정 자료에 대한 검색을 할 때 활용하는 함수이다. WHERE 이후에 대문자 A 로 시작하는 크기에 상관 없이 모든 문자를 검색하도록 한다.

② where GNL_NM_CD in (select generic_code from code.drug_master_ra)

- 서브쿼리(subquery) : WHERE, FROM, HAVING절에 다른 SQL문장을 포함하는 형태이다. 서브쿼리를 사용하면 불필요한 테이블을 생성하지 않고도 바로 원하는 테이블을 생성 할 수 있기 때문에 코드를 간소화 할 수 있다. 서브쿼리 사용 시, 문장을 괄호로 둘러싸야 하며, ORDER BY 절은 포함 할 수 없다.

(예)



③ drug_ra_use = sum(ra_drug_t30, ra_drug_t53)

if drug_ra_use > 0 then DMARD=1

⇒ 류마티스 치료제(DMARD)약물을 처방받은 환자를 구별하기 위하여 T30 T53 테이블의 DMARD 약물의 총사용량을 모두 합하여 DMARD약물 변수를 생성함. 만약 약물 총사용량 변수값이 0보다 크면 이를 DMARD를 처방받은 환자로 볼 수 있다

□ 류마티스관절염 환자 중 DMARD 사용 환자 선정

위에서 생성한 RA t2040_ra_dmard 테이블에서 4월부터 6월까지 dmard를 사용한 명세서만 선택하고, 해당 기간 중 처음으로 dmard 처방을 받은 요양개시일자를 기준 일자로 정의한다.

```
/* 2015년 RA 환자 중 4-6월 기간 중 처음으로 DMARD 처방이 확인 된 환자 */
data index;
  set T2040_ra_dmard;
  if dmard=1;
  recu_fr_dt = input(recu_fr_dd, yymmdd10.); /*숫자형태의 변수를 날짜형태로 바꿔주는 과정*/
  if 4 <= month(recu_fr_dt) <= 6;
  format recu_fr_dt yymmdd10. ;
run;

proc sort data=index;
  by jid recu_fr_dt; /*환자 별로 요양개시일자 순서대로 정렬 */
run;

/* cohort 정의, index date 정의 */
data cohort; /* => eligible cohort */
  set index;
  by jid;
  if first.jid; /* 개인식별대체키를 기준으로 처음으로 dmard를 처방받은 명세서만 선택하고
               요양개시일자를 index date로 정의*/
  rename recu_fr_dt=index_date;
  keep jid sex_tp_cd pat_age recu_fr_dt ra dmard;
run;
```

위의 프로그램 결과 생성되는 테이블은 다음과 같다.

	JID	SEX_TP_CD	PAT_AGE	ra	DMARD	index_date
1	1004278	2	75	1	1	2015-06-26
2	1012241	2	57	1	1	2015-04-17
3	101568	2	65	1	1	2015-04-02
4	1026937	1	42	1	1	2015-06-04
5	105498	2	44	1	1	2015-05-11
6	1055840	2	74	1	1	2015-06-26
7	1059694	2	40	1	1	2015-04-02
8	1063514	2	54	1	1	2015-05-07
9	1068210	2	53	1	1	2015-04-20
10	1077186	2	52	1	1	2015-04-15
11	1078690	1	34	1	1	2015-04-03
12	1080463	2	56	1	1	2015-05-19

※ 프로그램 Tip

① recu_fr_dt = input(recu_fr_dd, yymmdd10.)

format recu_fr_dt yymmdd10.

⇒ STEP1. input 함수로 값을 읽어 들어 날짜포맷으로 정의

STEP2. format을 이용해 출력할 포맷을 지정

FORMAT 형태	출력 형태
MMDDYY8.	07-04-15

DDMMYY10.	04-07-2015
DATE9.	04JUL2015
MONYY5.	JUL2015

- SAS 날짜 함수

사용법	기능/참고사항
mdy(월, 일, 연도)	지정한 날짜를 sas날짜로 변환, 1960/01/01~
today() / date()	오늘 날짜를 sas 날짜로 변환
time()	현재 시간을 sas 시간(초)으로 변환(0~86400), 자정 12시~
day(sas날짜)	입력한 sas 날짜의 '일'반환 (1~31)
qtr(sas날짜)	입력한 sas 날짜의 '분기'반환 (1~4)
weekday(sas날짜)	입력한 sas 날짜의 '요일'반환 (일=1, 월=2, ..., 토=7)
month(sas날짜)	입력한 sas 날짜의 '월'반환 (1~12)
year(sas날짜)	입력한 sas 날짜의 '연도'반환
intnx('interval', from, increment, '<alignment>')	from에서 increment만큼 구간을 이동한 값 반환 interval : year, semiyear, month day, qtr, week, weekday, tenday, dtweek, demonth, semimonth, hour, minute, second alignment : 'B' beginning, 'M' middle, 'S' sameday 기간의 초, 중, 말 설정 가능

③ if 4 <= month(recu_fr_dt) <= 6

⇒ 4월부터 6월까지 명세서만 선택

□ 결과 발생 정의

T40 테이블에서 고관절 골절 상병코드를 이용하여 고관절 골절 발생을 정의한다. 이때 고관절 골절 발생은 기준일자 이후 발생하는 경우만으로 정의해야 하므로, 우선 정의된 기준일자를 T2040_ra_dmard 테이블에 조인하고, 기준일자 이후에 발생하는 고관절 골절만 선택한다.

```
/*결과발생정의 - 고관절 골절 발생 */
/*T2040_ra_dmard테이블에 기준일자를 left join*/

proc sql;
  create table t40_1 as
  select a.*, b.index_date
  from t2040_ra_dmard as a
  left join cohort as b
  on (a.jid=b.jid); /*개인식별대체키를 기준으로 통합*/
quit;

data t40_2;
  set t40_1;
  /*기준일자 + 6개월을 followup date로 정의*/
  followup_date=intnx('month', index_date, 6, 'sameday');
  if index_date <= input(recu_fr_dd, yymmdd10.) <= followup_date;
```

```

/*T40테이블에서 상병코드 중 고관절 골절 발생 변수 생성*/
if substr(SICK_CD,1,4) in ('S720','S721','S722') then fracture=1;
format followup_date yymmdd10.;
run;

proc sql;
create table t40_3 as
select jid, max(fracture) as fracture /*고관절 골절 발생*/
from t40_2
group by jid;
quit;

proc sql;
create table cohort1 as
select a.*, b.*
from cohort as a
left join t40_3 as b
on (a.jid=b.jid);/*개인식별대체키를 기준으로 통합*/
quit;

```

위의 프로그램 결과 생성되는 t40_2, cohort1테이블은 다음과 같다.

t40_2

	SPEC_ID_SNO	ra	DMARD	index_date	followup_date	fracture	JID	DMD_TP_CD
1	13603405	1	1	2015-06-26	2015-12-26	.	1004278	0
2	13603405	1	1	2015-06-26	2015-12-26	.	1004278	0
3	13603405	1	1	2015-06-26	2015-12-26	.	1004278	0
4	13603405	1	1	2015-06-26	2015-12-26	.	1004278	0
5	15985410	1	.	2015-06-26	2015-12-26	.	1004278	0
6	15985411	1	1	2015-06-26	2015-12-26	.	1004278	0
7	15985412	1	1	2015-06-26	2015-12-26	.	1004278	0
8	15985410	1	.	2015-06-26	2015-12-26	.	1004278	0
9	15985411	1	1	2015-06-26	2015-12-26	.	1004278	0
10	15985410	1	.	2015-06-26	2015-12-26	.	1004278	0
11	15985412	1	1	2015-06-26	2015-12-26	.	1004278	0

cohort1

	JID	SEX_TP_CD	PAT_AGE	ra	DMARD	index_date	fracture
1	1004278	2	75	1	1	2015-06-26	.
2	1012241	2	57	1	1	2015-04-17	.
3	101568	2	65	1	1	2015-04-02	.
4	1026937	1	42	1	1	2015-06-04	.
5	105498	2	44	1	1	2015-05-11	.
6	1055840	2	74	1	1	2015-06-26	.
7	1059694	2	40	1	1	2015-04-02	.
8	1063514	2	54	1	1	2015-05-07	.
9	1068210	2	53	1	1	2015-04-20	.
10	1077186	2	52	1	1	2015-04-15	.
11	1078690	1	34	1	1	2015-04-03	.

참고문헌

- 건강보험 청구자료 분석 매뉴얼. 건강보험심사평가원. 2015.
- 배은영, 김윤희, 이태진, 이혜재, 장은진. 의료기술 경제성 및 급여적정성 평가 기준 개선 연구. 건강보험심사평가원. 2015.
- 장은진, 안정훈, 정선영, 황진섭, 이자연, 심정임. 측정된 교란요인을 고려한 성과분석 방법. 한국보건의료연구원. 2013.
- 황진섭, 신상진, 김종희, 오성희, 강현경, 박승희, 장은진, 최지은, 현민경. 국내 보건의료 이차자료원 활용. 한국보건의료연구원. 2013.
- 건강보험심사평가원 홈페이지
<http://www.hira.or.kr>
- 건강보험심사평가원 보건의료빅데이터개방시스템 홈페이지
<http://opendata.hira.or.kr>
- Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015;12(10): e1001885. doi:10.1371/journal.pmed.1001885.
- Effects of disease-modifying antirheumatic drugs on nonvertebral fracture risk in rheumatoid arthritis: a population-based cohort study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2012;27(4):789-96. Epub 2011/12/14.
- Hatch EE, Palmer JR, Titus-Ernstoff L, Noller KL, Kaufman RH, Mittendorf R, et al. Cancer risk in women exposed to diethylstilbestrol in utero. Jama. 1998;280(7):630-4.
- Herbst AL, Ulfelder H, Poskanzer DC, Longo LD. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. 1971. American journal of obstetrics and gynecology. 1999;181(6):1574-5.
- Hoover RN, Hyer M, Pfeiffer RM, Adam E, Bond B, Cheville AL, et al. Adverse health outcomes in women exposed in utero to diethylstilbestrol. The New England journal of medicine. 2011;365(14):1304-14.
- Iron, K & Manuel, D. Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data. Online. ICES. 2007.

부록 1. RECORD 체크리스트

<부록1 표1> RECORD 체크리스트

Item No	STROBE Items	RECORD Items
Title and abstract	(a) Indicate the study's design with a commonly used term in the title or the abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included
	(b) Provide in the abstract an informative and balanced summary of what was done and what was found	RECORD 1.2: If applicable, the geographic region and time frame within which the study took place should be reported in the title or abstract RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract
Introduction		
Background / rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	(a) Cohort study: Give the eligibility criteria and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study: Give the eligibility criteria and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross-sectional study: Give the eligibility criteria and the sources and methods of selection of participants	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided
	(b) Cohort study: For matched studies, give matching criteria and number of exposed and unexposed. Case-control study: For matched studies, give matching criteria and	RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data

Item No		STROBE Items	RECORD Items
		the number of controls per case	linkage process, including the number of individuals with linked data at each stage
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided
Data sources / measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	
		(b) Describe any methods used to examine subgroups and interactions	
		(c) Explain how missing data were addressed	
		(d) Cohort study: If applicable, explain how loss to follow-up was addressed Case-control study: If applicable, explain how matching of cases and controls was addressed Cross-sectional study: If applicable, describe analytical methods taking account of sampling strategy	
		(e) Describe any sensitivity analyses	
Data access and cleaning methods		N/A	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population
Linkage		N/A	RECORD 12.2: Authors should provide information on the data cleaning methods used in the study RECORD 12.3: State whether the study included person-level,

Item No		STROBE Items	RECORD Items
			institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided
Result			
Participants	13	<p>(a) Report the numbers of individuals at each stage of the study (e.g. numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed)</p> <p>(b) Give reasons for nonparticipation at each stage</p> <p>(c) Consider use of a flow diagram</p>	RECORD 13.1: Describe in detail the selection of the persons included in the study (i.e. study population selection), including filtering based on data quality, data availability, and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram
Descriptive data	14	<p>(a) Give characteristics of study participants(e.g. demographic, clinical, and social) and information on exposures and potential confounders</p> <p>(b) Indicate the number of participants with missing data for each variable of interest</p> <p>(c) Cohort study: summarise follow-up time (e.g. average and total amount)</p>	
Outcome data	15	<p>Cohort study: Report numbers of outcome events or summary measures over time</p> <p>Case-control study: Report numbers in each exposure category or summary measures of exposure</p> <p>Cross-sectional study: Report numbers of outcome events or summary measures</p>	
Main results	16	<p>(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g. 95% confidence interval). Make clear which confounders were adjusted for and why they were included</p> <p>(b) Report category boundaries when continuous variables were categorized</p> <p>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</p>	
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions and sensitivity analyses	
Discussion			

	Item No	STROBE Items	RECORD Items
Key results	18	Summarise key results with reference to study objectives	
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	
Generalisability	21	Discuss the generalisability (external validity) of the study results	
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code
Accessibility of protocol, raw data, and programming code		N/A	

부록 2. 명세서 현황 분석

<부록2 표1> 요양종별에 대한 서식코드별 비중

종별	서식구분									
	보건기관외래의과(081)		보건기관외래치과(082)		보건기관외래한방(083)		정신과낮병동(091)		정신과입원(101)	
	건수	백분율	건수	백분율	건수	백분율	건수	백분율	건수	백분율
종합전문병원, 상급종합병원	—	—	—	—	—	—	60	0.00	416	0.00
종합병원	—	—	—	—	—	—	214	0.00	31,114	0.00
병원	—	—	—	—	—	—	24,073	0.00	528,263	0.06
요양병원	—	—	—	—	—	—	237	0.00	28,732	0.00
의원	—	—	—	—	—	—	13,101	0.00	31,521	0.00
치과병원	—	—	—	—	—	—	—	—	—	—
치과의원	—	—	—	—	—	—	—	—	—	—
조산원	—	—	—	—	—	—	—	—	—	—
보건소	4,553,164	0.49	169,712	0.02	867,842	0.09	—	—	—	—
보건지소	2,903,431	0.31	67,981	0.01	681,355	0.07	—	—	—	—
보건진료소	4,419,867	0.48	—	—	—	—	—	—	—	—
보건의료원	—	—	—	—	—	—	—	—	—	—
한방병원	—	—	—	—	—	—	—	—	—	—
한의원	—	—	—	—	—	—	—	—	—	—
계	11,876,462	1.28	237,693	0.03	1,617,178	0.16	37,685	0.00	620,046	0.06

<부록2 표2> 요양종별에 대한 서식코드별 비중

종별	서식구분									
	의과입원(021)		의과외래(031)		치과입원(041)		치과외래(051)		조산원입원(061)	
	건수	백분율	건수	백분율	건수	백분율	건수	백분율	건수	백분율
종합전문병원, 상급종합병원	2,199,073	0.24	37,559,164	4.07	9,879	0.00	699,154	0.08	—	—
종합병원	3,560,164	0.39	64,775,456	7.03	13,899	0.00	1,220,121	0.13	—	—
병원	3,806,971	0.41	64,912,956	7.04	4,900	0.00	110,500	0.01	—	—
요양병원	2,805,900	0.30	2,967,853	0.32	1,988	0.00	6,764	0.00	—	—
의원	1,756,505	0.19	538,232,626	58.39	—	—	—	—	—	—
치과병원	—	—	3,846	0.00	5,385	0.00	3,611,024	0.39	—	—
치과의원	—	—	—	—	34	0.00	65,591,920	7.12	—	—
조산원	—	—	—	—	—	—	—	—	1,228	0.00
보건소	—	—	—	—	—	—	—	—	—	—
보건지소	—	—	—	—	—	—	—	—	—	—
보건진료소	—	—	—	—	—	—	—	—	—	—
보건의료원	3,143	0.00	543,094	0.06	—	—	11,219	0.00	—	—
한방병원	141,327	0.02	290,654	0.03	—	—	2,377	0.00	—	—
한의원	—	—	—	—	—	—	—	—	—	—
계	14,273,083	1.55	709,285,649	76.94	36,085	0.00	71,253,079	7.73	1,228	0.00

<부록2 표3> 요양종별에 대한 서식코드별 비중

종별	서식구분							
	정신과외래(111)		한방입원(121)		한방외래(131)		의료급여혈액투석정액(151)	
	건수	백분율	건수	백분율	건수	백분율	건수	백분율
종합전문병원, 상급종합병원	65,252	0.01	—	—	214	0.00	62,863	0.01
종합병원	364,890	0.04	5,253	0.00	66,941	0.01	557,375	0.06
병원	555,572	0.06	51,346	0.01	143,868	0.02	131,828	0.01
요양병원	28,928	0.00	1,129,768	0.12	313,191	0.03	65,276	0.01
의원	1,151,982	0.12	—	—	—	—	867,719	0.09
치과병원	—	—	—	—	24,480	0.00	—	—
치과의원	—	—	—	—	—	—	—	—
조산원	—	—	—	—	—	—	—	—
보건소	—	—	—	—	—	—	—	—
보건지소	—	—	—	—	—	—	—	—
보건진료소	—	—	—	—	—	—	—	—
보건의료원	1,598	0.00	—	—	33,837	0.00	—	—
한방병원	—	—	213,361	0.02	2,963,801	0.32	—	—
한의원	—	—	18,970	0.00	103,816,092	11.26	—	—
계	2,168,222	0.23	1,418,698	0.15	107,362,424	11.64	1,685,061	0.18

<부록2 표4> 서식 구분

서식구분코드	서식구분 코드명	빈도	백분율(%)
021	의과입원	14,273,083	1.55
031	의과외래	709,285,649	76.95
041	치과입원	36,085	0.00
051	치과외래	71,253,079	7.73
061	조산원입원	1,228	0.00
081	보건기관외래의과	11,876,462	1.29
082	보건기관외래치과	237,693	0.03
083	보건기관외래한방	1,549,197	0.17
091	정신과낮병동	37,685	0.00
101	정신과입원	620,046	0.07
111	정신과외래	2,168,222	0.24
121	한방입원	1,418,698	0.15
131	한방외래	107,362,424	11.65
151	의료급여혈액투석정액	1,685,061	0.18
전체		921,804,612	100

<부록2 표5> 청구구분별 비중

청구구분코드	청구구분코드명	건수	비중
0	일반청구	916,546,874	99.43
1	보완청구	460,870	0.05
2	추가청구	321,895	0.03
3	분리청구	4,474,928	0.49
9	신포괄채청구구분	45	0.00
계		921,804,612	100.00

<부록2 표6> 청구형태별 비중

청구형태코드	청구형태코드명	건수	비중
0	서면	92,627	0.01
1	전산매체(디스켓,CD등)	4,244,000	0.46
2	전자문서(EDI,청구포탈)	916,354,790	99.41
3	DRG	1,113,195	0.12
계		921,804,612	100.00

<부록2 표7> 수진자구분별 비중

수진자구분코드	수진자구분코드명	건수	비중
1	일반	912,456,975	98.99
2	신생아	658,711	0.07
3	행려	86,662	0.01
4	외국인	8,597,021	0.93
5	외국인(신생아)	1,235	0.00
6	주민등록번호 오류	4,008	0.00
계		921,804,612	100.00

<부록2 표8> 보험자종별 비중

보험자종별코드	보험자종별코드명	건수	비중
4	건강보험	869,825,910	94.36
5	의료급여	50,387,440	5.47
7	보훈	1,591,262	0.17
계		921,804,612	100.00

<부록2 표9> 상병분류별 비중

상병분류코드	상병분류코드명	건수	비중
\$	결측	1,113,195	0.07
1	주상병	920,691,417	59.37
2	부상병	611,913,781	39.46
3	배제상병	17,023,542	1.10
계		1,550,741,935	100.00

<부록2 표10> 명세서별 상병 개수 분포

상병순위	명세서건수	백분율	누적백분율
주상병	305,793,406	33.17	33.17
주상병+1부상병	215,228,580	23.35	56.52
주상병+2부상병	160,105,407	17.37	73.89
주상병+3부상병	101,510,974	11.01	84.90
주상병+4부상병	58,490,174	6.35	91.25
주상병+5부상병	31,887,661	3.46	94.71
주상병+6부상병	17,689,032	1.92	96.63
주상병+7부상병	10,377,051	1.13	97.75
주상병+8부상병	6,418,171	0.70	98.45
주상병+9부상병	4,180,088	0.45	98.90
주상병+10부상병	2,823,937	0.31	99.21
주상병+11부상병	2,390,910	0.26	99.47
주상병+12부상병	1,208,732	0.13	99.60

주상병+13부상병	886,589	0.10	99.69
주상병+14부상병	653,841	0.07	99.77
주상병+15부상병	486,934	0.05	99.82
주상병+16부상병	370,405	0.04	99.86
주상병+17부상병	280,412	0.03	99.89
주상병+18부상병	214,195	0.02	99.91
주상병+19부상병	170,494	0.02	99.93
주상병+20부상병	128,929	0.01	99.94
주상병+21부상병	100,104	0.01	99.96
주상병+22부상병	82,335	0.01	99.96
주상병+23부상병	63,490	0.01	99.97
주상병+24부상병	51,748	0.01	99.98
주상병+25부상병	43,024	0.00	99.98
주상병+26부상병	33,535	0.00	99.99
주상병+27부상병	28,363	0.00	99.99
주상병+28부상병	23,539	0.00	99.99
주상병+29부상병	67,880	0.01	100.00
주상병+30부상병	3,527	0.00	100.00
주상병+31부상병	2,507	0.00	100.00
주상병+32부상병	1,905	0.00	100.00
주상병+33부상병	1,567	0.00	100.00
주상병+34부상병	1,250	0.00	100.00
주상병+35부상병	867	0.00	100.00
주상병+36부상병	828	0.00	100.00
주상병+37부상병	673	0.00	100.00
주상병+38부상병	463	0.00	100.00
주상병+39부상병	379	0.00	100.00
주상병+40부상병	293	0.00	100.00
주상병+41부상병	111	0.00	100.00
주상병+42부상병	94	0.00	100.00
주상병+43부상병	39	0.00	100.00
주상병+44부상병	17	0.00	100.00
주상병+45부상병	83	0.00	100.00
주상병+46부상병	26	0.00	100.00
주상병+47부상병	7	0.00	100.00
주상병+48부상병	8	0.00	100.00
주상병+49부상병	8	0.00	100.00
주상병+50부상병	8	0.00	100.00
주상병+52부상병	2	0.00	100.00
주상병+55부상병	6	0.00	100.00
주상병+56부상병	3	0.00	100.00
주상병+65부상병	1	0.00	100.00
계	921,804,612	100.00	

<부록2 표11> 전체 암 상병정의 개수에 따른 건수(전체 암)

상병순위	명세서건수	백분율	누적백분율
주상병	10,983,797	82.49	82.49
주상병+부상병1	1,114,374	8.37	90.86
주상병+부상병2	408,959	3.07	93.93

주상병+부상병3	251,687	1.89	95.82
주상병+부상병4	159,716	1.20	97.02
주상병+부상병5	106,462	0.80	97.82
주상병+부상병6	71,657	0.54	98.36
주상병+부상병7	50,074	0.38	98.74
주상병+부상병8	36,493	0.27	99.01
주상병+부상병9	27,500	0.21	99.22
주상병+부상병10	20,946	0.16	99.37
주상병+부상병11	16,568	0.12	99.50
주상병+부상병12	12,572	0.09	99.59
주상병+부상병13	10,619	0.08	99.67
주상병+부상병14	8,500	0.06	99.74
주상병+부상병15	6,621	0.05	99.79
주상병+부상병16	5,556	0.04	99.83
주상병+부상병17	4,515	0.03	99.86
주상병+부상병18	3,829	0.03	99.89
주상병+부상병19	2,894	0.02	99.91
주상병+부상병20	2,501	0.02	99.93
주상병+부상병21	1,960	0.01	99.95
주상병+부상병22	1,561	0.01	99.96
주상병+부상병23	1,315	0.01	99.97
주상병+부상병24	1,019	0.01	99.98
주상병+부상병25	794	0.01	99.98
주상병+부상병26	695	0.01	99.99
주상병+부상병27	575	0.00	99.99
주상병+부상병28	409	0.00	99.99
주상병+부상병29	307	0.00	100.00
주상병+부상병30	178	0.00	100.00
주상병+부상병31	91	0.00	100.00
주상병+부상병32	63	0.00	100.00
주상병+부상병33	41	0.00	100.00
주상병+부상병34	27	0.00	100.00
주상병+부상병35	25	0.00	100.00
주상병+부상병36	11	0.00	100.00
주상병+부상병37	7	0.00	100.00
주상병+부상병38	4	0.00	100.00
주상병+부상병39	5	0.00	100.00
주상병+부상병40	1	0.00	100.00
주상병+부상병41	2	0.00	100.00
주상병+부상병42	12	0.00	100.00
주상병+부상병43	3	0.00	100.00
주상병+부상병44	11	0.00	100.00
주상병+부상병45	1	0.00	100.00
주상병+부상병47	1	0.00	100.00
계	13,314,958	100.00	

<부록2 표12> 상병정의 개수에 따른 건수(당노병)

상병순위	명세서건수	백분율	누적백분율
주상병	19,904,957	45.73	45.73

주상병+부상병1	10,732,138	24.66	70.39
주상병+부상병2	5,001,923	11.49	81.89
주상병+부상병3	2,927,179	6.73	88.61
주상병+부상병4	1,760,885	4.05	92.66
주상병+부상병5	1,091,267	2.51	95.16
주상병+부상병6	678,602	1.56	96.72
주상병+부상병7	434,606	1.00	97.72
주상병+부상병8	287,432	0.66	98.38
주상병+부상병9	196,491	0.45	98.83
주상병+부상병10	137,671	0.32	99.15
주상병+부상병11	98,190	0.23	99.38
주상병+부상병12	66,686	0.15	99.53
주상병+부상병13	50,587	0.12	99.65
주상병+부상병14	36,693	0.08	99.73
주상병+부상병15	27,100	0.06	99.79
주상병+부상병16	20,436	0.05	99.84
주상병+부상병17	16,179	0.04	99.88
주상병+부상병18	12,276	0.03	99.90
주상병+부상병19	9,647	0.02	99.93
주상병+부상병20	7,458	0.02	99.94
주상병+부상병21	5,457	0.01	99.96
주상병+부상병22	4,462	0.01	99.97
주상병+부상병23	3,769	0.01	99.97
주상병+부상병24	2,808	0.01	99.98
주상병+부상병25	2,542	0.01	99.99
주상병+부상병26	2,001	0.00	99.99
주상병+부상병27	1,495	0.00	100.00
주상병+부상병28	899	0.00	100.00
주상병+부상병29	776	0.00	100.00
주상병+부상병30	136	0.00	100.00
주상병+부상병31	94	0.00	100.00
주상병+부상병32	36	0.00	100.00
주상병+부상병33	58	0.00	100.00
주상병+부상병34	33	0.00	100.00
주상병+부상병35	22	0.00	100.00
주상병+부상병36	23	0.00	100.00
주상병+부상병37	67	0.00	100.00
주상병+부상병38	18	0.00	100.00
주상병+부상병39	2	0.00	100.00
주상병+부상병40	1	0.00	100.00
주상병+부상병42	6	0.00	100.00
주상병+부상병43	2	0.00	100.00
주상병+부상병48	1	0.00	100.00
계	43,523,111	100.00	

<부록2 표13> 상병정의 개수에 따른 건수(고혈압)

상병순위	명세서건수	백분율	누적백분율
주상병	45,917,523	57.59	57.59
주상병+부상병1	15,254,404	19.13	76.72

주상병+부상병2	8,114,360	10.18	86.90
주상병+부상병3	4,505,327	5.65	92.55
주상병+부상병4	2,462,643	3.09	95.64
주상병+부상병5	1,377,951	1.73	97.37
주상병+부상병6	784,262	0.98	98.35
주상병+부상병7	464,579	0.58	98.93
주상병+부상병8	287,551	0.36	99.29
주상병+부상병9	181,302	0.23	99.52
주상병+부상병10	120,738	0.15	99.67
주상병+부상병11	77,179	0.10	99.77
주상병+부상병12	51,855	0.07	99.83
주상병+부상병13	35,407	0.04	99.88
주상병+부상병14	25,048	0.03	99.91
주상병+부상병15	17,064	0.02	99.93
주상병+부상병16	13,566	0.02	99.95
주상병+부상병17	10,267	0.01	99.96
주상병+부상병18	7,885	0.01	99.97
주상병+부상병19	6,159	0.01	99.98
주상병+부상병20	4,790	0.01	99.98
주상병+부상병21	3,147	0.00	99.99
주상병+부상병22	2,845	0.00	99.99
주상병+부상병23	1,809	0.00	99.99
주상병+부상병24	1,489	0.00	100.00
주상병+부상병25	933	0.00	100.00
주상병+부상병26	871	0.00	100.00
주상병+부상병27	940	0.00	100.00
주상병+부상병28	319	0.00	100.00
주상병+부상병29	636	0.00	100.00
주상병+부상병30	37	0.00	100.00
주상병+부상병31	13	0.00	100.00
주상병+부상병32	14	0.00	100.00
주상병+부상병33	10	0.00	100.00
주상병+부상병34	11	0.00	100.00
주상병+부상병35	8	0.00	100.00
주상병+부상병36	10	0.00	100.00
주상병+부상병37	6	0.00	100.00
주상병+부상병38	1	0.00	100.00
주상병+부상병41	3	0.00	100.00
주상병+부상병42	1	0.00	100.00
계	79,732,963	100.00	

<부록2 표14> 상병정의 개수에 따른 건수(천식)

상병순위	명세서건수	백분율	누적백분율
주상병	5,518,209	20.72	20.72
주상병+부상병1	6,101,945	22.92	43.64
주상병+부상병2	5,424,756	20.37	64.02
주상병+부상병3	3,966,738	14.90	78.91
주상병+부상병4	2,402,831	9.02	87.94
주상병+부상병5	1,304,642	4.90	92.84

주상병+부상병6	701,535	2.63	95.47
주상병+부상병7	411,379	1.55	97.02
주상병+부상병8	248,501	0.93	97.95
주상병+부상병9	162,064	0.61	98.56
주상병+부상병10	109,868	0.41	98.97
주상병+부상병11	73,973	0.28	99.25
주상병+부상병12	46,445	0.17	99.42
주상병+부상병13	34,720	0.13	99.56
주상병+부상병14	24,326	0.09	99.65
주상병+부상병15	19,556	0.07	99.72
주상병+부상병16	15,578	0.06	99.78
주상병+부상병17	11,958	0.04	99.82
주상병+부상병18	9,632	0.04	99.86
주상병+부상병19	8,763	0.03	99.89
주상병+부상병20	6,294	0.02	99.92
주상병+부상병21	4,873	0.02	99.93
주상병+부상병22	4,354	0.02	99.95
주상병+부상병23	3,577	0.01	99.96
주상병+부상병24	2,211	0.01	99.97
주상병+부상병25	2,100	0.01	99.98
주상병+부상병26	1,328	0.00	99.99
주상병+부상병27	1,102	0.00	99.99
주상병+부상병28	1,146	0.00	99.99
주상병+부상병29	1,001	0.00	100.00
주상병+부상병30	179	0.00	100.00
주상병+부상병31	64	0.00	100.00
주상병+부상병32	131	0.00	100.00
주상병+부상병33	37	0.00	100.00
주상병+부상병34	150	0.00	100.00
주상병+부상병35	3	0.00	100.00
주상병+부상병36	31	0.00	100.00
주상병+부상병37	6	0.00	100.00
주상병+부상병38	31	0.00	100.00
주상병+부상병39	2	0.00	100.00
주상병+부상병41	1	0.00	100.00
계	26,626,040	100.00	

<부록2 표15> 상병정의 개수에 따른 건수(관절증)

상병순위	명세서건수	백분율	누적백분율
주상병	23,345,514	47.37	47.37
주상병+부상병1	11,347,359	23.02	70.40
주상병+부상병2	6,080,846	12.34	82.73
주상병+부상병3	3,374,129	6.85	89.58
주상병+부상병4	1,935,040	3.93	93.51
주상병+부상병5	1,140,072	2.31	95.82
주상병+부상병6	686,029	1.39	97.21
주상병+부상병7	428,283	0.87	98.08
주상병+부상병8	284,530	0.58	98.66
주상병+부상병9	187,683	0.38	99.04

주상병+부상병10	130,058	0.26	99.30
주상병+부상병11	93,252	0.19	99.49
주상병+부상병12	63,026	0.13	99.62
주상병+부상병13	45,418	0.09	99.71
주상병+부상병14	32,836	0.07	99.78
주상병+부상병15	25,721	0.05	99.83
주상병+부상병16	20,143	0.04	99.87
주상병+부상병17	15,131	0.03	99.90
주상병+부상병18	11,367	0.02	99.93
주상병+부상병19	8,164	0.02	99.94
주상병+부상병20	6,531	0.01	99.96
주상병+부상병21	5,133	0.01	99.97
주상병+부상병22	4,470	0.01	99.98
주상병+부상병23	3,205	0.01	99.98
주상병+부상병24	2,834	0.01	99.99
주상병+부상병25	1,622	0.00	99.99
주상병+부상병26	1,459	0.00	99.99
주상병+부상병27	1,015	0.00	100.00
주상병+부상병28	859	0.00	100.00
주상병+부상병29	660	0.00	100.00
주상병+부상병30	164	0.00	100.00
주상병+부상병31	99	0.00	100.00
주상병+부상병32	102	0.00	100.00
주상병+부상병33	61	0.00	100.00
주상병+부상병34	42	0.00	100.00
주상병+부상병35	13	0.00	100.00
주상병+부상병36	21	0.00	100.00
주상병+부상병37	34	0.00	100.00
주상병+부상병38	8	0.00	100.00
주상병+부상병39	2	0.00	100.00
주상병+부상병40	3	0.00	100.00
주상병+부상병41	2	0.00	100.00
주상병+부상병42	1	0.00	100.00
계	49,282,941	100.00	

<부록2 표16> 상병정의 개수에 따른 건수(류마티스 관절염)

상병순위	명세서건수	백분율	누적백분율
주상병	1,385,974	36.25	36.25
주상병+부상병1	663,620	17.36	53.60
주상병+부상병2	510,866	13.36	66.96
주상병+부상병3	370,453	9.69	76.65
주상병+부상병4	265,572	6.95	83.60
주상병+부상병5	179,620	4.70	88.30
주상병+부상병6	121,349	3.17	91.47
주상병+부상병7	83,624	2.19	93.66
주상병+부상병8	61,749	1.61	95.27
주상병+부상병9	44,172	1.16	96.43
주상병+부상병10	32,849	0.86	97.29
주상병+부상병11	23,993	0.63	97.91

주상병+부상병12	17,061	0.45	98.36
주상병+부상병13	13,233	0.35	98.71
주상병+부상병14	10,420	0.27	98.98
주상병+부상병15	8,205	0.21	99.19
주상병+부상병16	6,245	0.16	99.36
주상병+부상병17	5,079	0.13	99.49
주상병+부상병18	4,090	0.11	99.60
주상병+부상병19	3,197	0.08	99.68
주상병+부상병20	2,456	0.06	99.74
주상병+부상병21	2,476	0.06	99.81
주상병+부상병22	1,743	0.05	99.85
주상병+부상병23	1,269	0.03	99.89
주상병+부상병24	1,151	0.03	99.92
주상병+부상병25	858	0.02	99.94
주상병+부상병26	583	0.02	99.96
주상병+부상병27	841	0.02	99.98
주상병+부상병28	440	0.01	99.99
주상병+부상병29	270	0.01	100.00
주상병+부상병30	53	0.00	100.00
주상병+부상병31	31	0.00	100.00
주상병+부상병32	19	0.00	100.00
주상병+부상병33	21	0.00	100.00
주상병+부상병34	12	0.00	100.00
주상병+부상병35	16	0.00	100.00
주상병+부상병36	3	0.00	100.00
주상병+부상병37	1	0.00	100.00
주상병+부상병38	3	0.00	100.00
주상병+부상병48	1	0.00	100.00
계	3,823,618	100.00	

<부록2 표17> 상병정의 개수에 따른 건수(신부전증)

상병순위	명세서건수	백분율	누적백분율
주상병	8,458,822	86.57	86.57
주상병+부상병1	424,892	4.35	90.92
주상병+부상병2	235,657	2.41	93.34
주상병+부상병3	180,481	1.85	95.18
주상병+부상병4	131,028	1.34	96.52
주상병+부상병5	92,494	0.95	97.47
주상병+부상병6	66,322	0.68	98.15
주상병+부상병7	47,185	0.48	98.63
주상병+부상병8	33,379	0.34	98.97
주상병+부상병9	24,054	0.25	99.22
주상병+부상병10	18,015	0.18	99.40
주상병+부상병11	14,095	0.14	99.55
주상병+부상병12	9,527	0.10	99.65
주상병+부상병13	7,218	0.07	99.72
주상병+부상병14	5,991	0.06	99.78
주상병+부상병15	4,581	0.05	99.83
주상병+부상병16	3,526	0.04	99.86
주상병+부상병17	3,059	0.03	99.90

주상병+부상병18	2,288	0.02	99.92
주상병+부상병19	2,092	0.02	99.94
주상병+부상병20	1,402	0.01	99.95
주상병+부상병21	1,080	0.01	99.97
주상병+부상병22	813	0.01	99.97
주상병+부상병23	648	0.01	99.98
주상병+부상병24	645	0.01	99.99
주상병+부상병25	362	0.00	99.99
주상병+부상병26	276	0.00	99.99
주상병+부상병27	241	0.00	100.00
주상병+부상병28	221	0.00	100.00
주상병+부상병29	68	0.00	100.00
주상병+부상병30	26	0.00	100.00
주상병+부상병31	18	0.00	100.00
주상병+부상병32	9	0.00	100.00
주상병+부상병33	3	0.00	100.00
주상병+부상병34	7	0.00	100.00
주상병+부상병35	4	0.00	100.00
주상병+부상병36	4	0.00	100.00
주상병+부상병37	2	0.00	100.00
주상병+부상병39	1	0.00	100.00
주상병+부상병44	1	0.00	100.00
계	9,770,537	100.00	

<부록2 표18> 상병정의 개수에 따른 건수(뇌혈관 질환)

상병순위	명세서건수	백분율	누적백분율
주상병	4,852,439	9.14	9.14
주상병+부상병1	18,491,859	34.85	43.99
주상병+부상병2	12,718,304	23.97	67.96
주상병+부상병3	6,797,978	12.81	80.77
주상병+부상병4	3,864,669	7.28	88.05
주상병+부상병5	2,319,987	4.37	92.42
주상병+부상병6	1,401,891	2.64	95.06
주상병+부상병7	866,760	1.63	96.70
주상병+부상병8	554,084	1.04	97.74
주상병+부상병9	363,904	0.69	98.43
주상병+부상병10	249,584	0.47	98.90
주상병+부상병11	166,424	0.31	99.21
주상병+부상병12	112,385	0.21	99.42
주상병+부상병13	79,097	0.15	99.57
주상병+부상병14	58,683	0.11	99.68
주상병+부상병15	41,711	0.08	99.76
주상병+부상병16	30,919	0.06	99.82
주상병+부상병17	23,655	0.04	99.86
주상병+부상병18	17,149	0.03	99.90
주상병+부상병19	13,535	0.03	99.92
주상병+부상병20	10,405	0.02	99.94
주상병+부상병21	7,490	0.01	99.96
주상병+부상병22	6,213	0.01	99.97

주상병+부상병23	4,453	0.01	99.98
주상병+부상병24	3,647	0.01	99.98
주상병+부상병25	2,725	0.01	99.99
주상병+부상병26	2,080	0.00	99.99
주상병+부상병27	1,605	0.00	99.99
주상병+부상병28	1,361	0.00	100.00
주상병+부상병29	788	0.00	100.00
주상병+부상병30	134	0.00	100.00
주상병+부상병31	87	0.00	100.00
주상병+부상병32	125	0.00	100.00
주상병+부상병33	58	0.00	100.00
주상병+부상병34	34	0.00	100.00
주상병+부상병35	43	0.00	100.00
주상병+부상병36	11	0.00	100.00
주상병+부상병37	15	0.00	100.00
주상병+부상병38	4	0.00	100.00
주상병+부상병39	5	0.00	100.00
주상병+부상병40	4	0.00	100.00
주상병+부상병41	6	0.00	100.00
주상병+부상병43	3	0.00	100.00
주상병+부상병44	1	0.00	100.00
주상병+부상병45	3	0.00	100.00
주상병+부상병46	1	0.00	100.00
주상병+부상병52	1	0.00	100.00
계	53,066,319	100.00	

<부록2 표19> 상병정의 개수에 따른 건수(이상지질혈증)

상병순위	명세서건수	백분율	누적백분율
주상병	4,852,439	9.14	9.14
주상병+부상병1	18,491,859	34.85	43.99
주상병+부상병2	12,718,304	23.97	67.96
주상병+부상병3	6,797,978	12.81	80.77
주상병+부상병4	3,864,669	7.28	88.05
주상병+부상병5	2,319,987	4.37	92.42
주상병+부상병6	1,401,891	2.64	95.06
주상병+부상병7	866,760	1.63	96.70
주상병+부상병8	554,084	1.04	97.74
주상병+부상병9	363,904	0.69	98.43
주상병+부상병10	249,584	0.47	98.90
주상병+부상병11	166,424	0.31	99.21
주상병+부상병12	112,385	0.21	99.42
주상병+부상병13	79,097	0.15	99.57
주상병+부상병14	58,683	0.11	99.68
주상병+부상병15	41,711	0.08	99.76
주상병+부상병16	30,919	0.06	99.82
주상병+부상병17	23,655	0.04	99.86
주상병+부상병18	17,149	0.03	99.90
주상병+부상병19	13,535	0.03	99.92
주상병+부상병20	10,405	0.02	99.94

주상병+부상병21	7,490	0.01	99.96
주상병+부상병22	6,213	0.01	99.97
주상병+부상병23	4,453	0.01	99.98
주상병+부상병24	3,647	0.01	99.98
주상병+부상병25	2,725	0.01	99.99
주상병+부상병26	2,080	0.00	99.99
주상병+부상병27	1,605	0.00	99.99
주상병+부상병28	1,361	0.00	100.00
주상병+부상병29	788	0.00	100.00
주상병+부상병30	134	0.00	100.00
주상병+부상병31	87	0.00	100.00
주상병+부상병32	125	0.00	100.00
주상병+부상병33	58	0.00	100.00
주상병+부상병34	34	0.00	100.00
주상병+부상병35	43	0.00	100.00
주상병+부상병36	11	0.00	100.00
주상병+부상병37	15	0.00	100.00
주상병+부상병38	4	0.00	100.00
주상병+부상병39	5	0.00	100.00
주상병+부상병40	4	0.00	100.00
주상병+부상병41	6	0.00	100.00
주상병+부상병43	3	0.00	100.00
주상병+부상병44	1	0.00	100.00
주상병+부상병45	3	0.00	100.00
주상병+부상병46	1	0.00	100.00
주상병+부상병52	1	0.00	100.00
계	53,066,319	100.00	

<부록2 표20> 상병정의 개수에 따른 건수(허혈성 심장질환)

상병순위	명세서건수	백분율	누적백분율
주상병	3,611,659	31.46	31.46
주상병+부상병1	2,394,577	20.86	52.31
주상병+부상병2	1,793,291	15.62	67.93
주상병+부상병3	1,228,225	10.70	78.63
주상병+부상병4	787,988	6.86	85.49
주상병+부상병5	505,183	4.40	89.89
주상병+부상병6	340,278	2.96	92.85
주상병+부상병7	231,750	2.02	94.87
주상병+부상병8	157,168	1.37	96.24
주상병+부상병9	111,594	0.97	97.21
주상병+부상병10	80,411	0.70	97.91
주상병+부상병11	58,789	0.51	98.42
주상병+부상병12	40,266	0.35	98.77
주상병+부상병13	31,021	0.27	99.04
주상병+부상병14	23,023	0.20	99.25
주상병+부상병15	18,539	0.16	99.41
주상병+부상병16	14,935	0.13	99.54
주상병+부상병17	12,601	0.11	99.65
주상병+부상병18	9,253	0.08	99.73

주상병+부상병19	6,448	0.06	99.78
주상병+부상병20	5,370	0.05	99.83
주상병+부상병21	4,149	0.04	99.87
주상병+부상병22	3,688	0.03	99.90
주상병+부상병23	3,069	0.03	99.93
주상병+부상병24	2,392	0.02	99.95
주상병+부상병25	1,724	0.02	99.96
주상병+부상병26	1,244	0.01	99.97
주상병+부상병27	995	0.01	99.98
주상병+부상병28	841	0.01	99.99
주상병+부상병29	605	0.01	99.99
주상병+부상병30	172	0.00	99.99
주상병+부상병31	179	0.00	100.00
주상병+부상병32	60	0.00	100.00
주상병+부상병33	75	0.00	100.00
주상병+부상병34	146	0.00	100.00
주상병+부상병35	42	0.00	100.00
주상병+부상병36	54	0.00	100.00
주상병+부상병37	35	0.00	100.00
주상병+부상병38	2	0.00	100.00
주상병+부상병39	8	0.00	100.00
주상병+부상병41	14	0.00	100.00
주상병+부상병42	11	0.00	100.00
계	11,481,874	100.00	